







# **Fluid Mechanics and Statistical Methods in Engineering**





UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

# Fluid Mechanics and Statistical Methods in Engineering

*By*

HUGH L. DRYDEN  
THEODORE VON KÁRMÁN  
ANTON A. KALINSKE  
THOMAS K. SHERWOOD  
SAMUEL S. WILKS  
WALTER A. SHEWHART  
LESLIE E. SIMON  
ROSCOE POUND



UNIVERSITY OF PENNSYLVANIA PRESS  
Philadelphia  
1941



UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

The Rôle of Transition from Laminar to  
Turbulent Flow in Fluid Mechanics

*By*

HUGH L. DRYDEN, PH.D.\*

INTRODUCTION

FLUID mechanics is one of the basic sciences which finds practical application in such diverse fields as aeronautics, hydraulic engineering, ventilation, chemical engineering, sanitary engineering, marine engineering, automotive engineering, railroad engineering, turbine design, and ballistics. In these many fields, flow relationships of such bewildering diversity are exhibited that at first sight the observed phenomena seem to have little in common. In reality the relationships are no more complicated than those involved in avoiding the failure of structures or structural parts under service conditions. The engineer soon learns to recognize certain characteristic types of behavior occurring in failures of a great variety of structures, and to study these types of failure under the simplest possible conditions. Thus, buckling is recognized as a type of failure characteristic of thin-walled structures or parts of structures, and a study of the buckling of a simple strut under axial loading aids in the understanding of the buckling of more complicated structures.

In a similar way it is possible to pick out certain characteristic flow phenomena which can be recognized in very complex flow patterns and in diverse types of apparatus and structures. The study of these characteristic phenomena under the simplest possible conditions aids greatly in the understanding of the flow of fluids in general. Some of these phenomena will be discussed in this paper.

\* Chief, Mechanics and Sound Division, National Bureau of Standards.

## TYPES OF FLOW AROUND A CYLINDER

The principal phenomena may be illustrated by the flow about a very long circular cylinder immersed in a fluid and subjected to flow normal to its axis at successively higher speeds. It is well known that the fields of flow about cylinders of various diameters in various fluids and at various speeds are similar if the ratio of the product of diameter  $D$  by speed  $U$  to the kinematic viscosity  $\nu$  is the same. This ratio  $UD/\nu$  is called the Reynolds Number, and is the only significant parameter provided that  $D$  is not comparable with the mean free path of the molecules of the fluid, that  $U$  is not comparable with the speed of sound in the fluid, and that the cylinder is not near the free surface of a liquid. For a cylinder one inch in diameter in air at normal pressure and at a temperature of 78°F, a Reynolds Number of 1 corresponds to a speed of 0.002 ft/sec, 100 to 0.2 ft/sec, 10,000 to 20 ft/sec, etc. For the same cylinder in water at 90°F a Reynolds Number of 1 corresponds to  $10^{-4}$  ft/sec, 100 to  $10^{-2}$  ft/sec, 10,000 to 1 ft/sec, etc. It is hardly possible to cover experimentally a range in Reynolds Number from 1 to 300,000 using one cylinder because of various practical difficulties, yet it may be helpful to consider an increasing Reynolds Number as corresponding to an increasing speed.

Fig. 1 is a concise summary of the known facts. Six general types of flow are distinguished and illustrated by sketches and graphs of pertinent measurements. The sketches for A, E, and F are schematic only, but the others are taken from published experimental results.

The six types of flow correspond to definite characteristics in the behavior of the total force exerted on the cylinder by the fluid stream. The force is usually expressed in terms of the drag coefficient,  $C_d$ , defined as the ratio of the force to the product of projected area (diameter times length) and velocity pressure (one-half the product of density by square of the speed). The variation of the drag coefficient with Reynolds Number is shown in Fig. 2 with logarithmic scales, the solid curve corresponding to a fluid stream of reasonably small initial turbulence and the dotted curve to a stream of large initial turbulence (1).<sup>\*</sup> The turbulence of the stream referred to here and elsewhere in the paper is a small-scale

<sup>\*</sup> A number in parentheses refers to the list of references at the end of the paper.

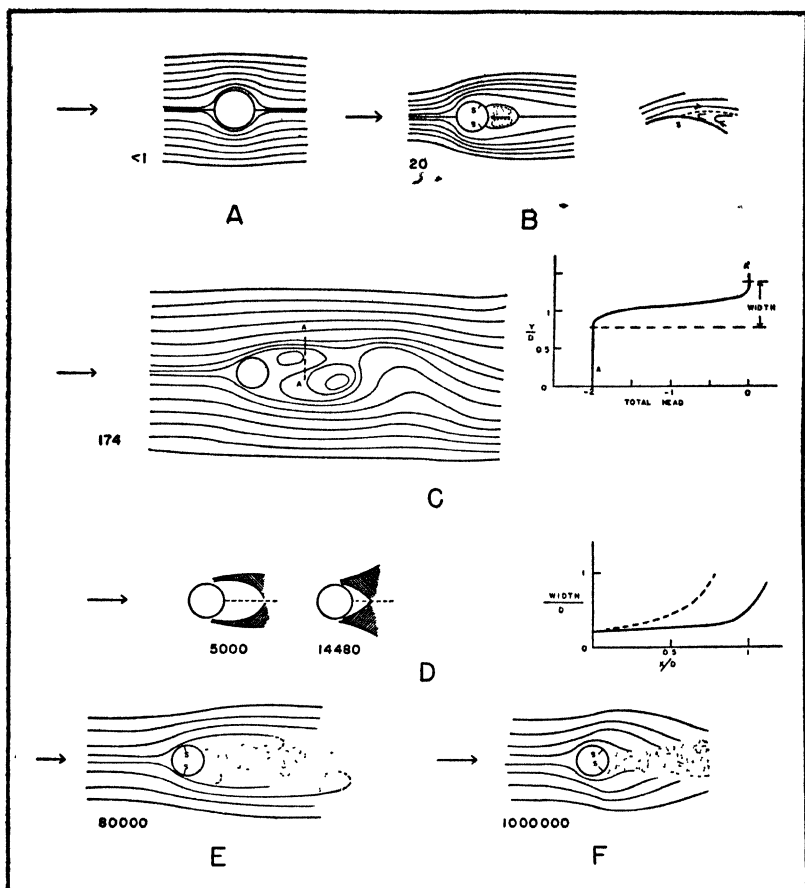


FIG. 1. Types of flow around an infinitely long circular cylinder.

A. At Reynolds Numbers less than 1.

B. At Reynolds Number of 20. S is the separation point of the laminar boundary layer.

C. At Reynolds Number of 174. One stage of the flow which varies with the time as eddies are shed alternately. Traverse with total head tube along AA' shown at the right, with width of free laminar layer indicated.

D. Boundaries of free layer at Reynolds Numbers of 5000 and 14480, and graphs of variation of width of free layer with distance X downstream. D is the diameter of the cylinder. The solid curve in the graph is for a Reynolds Number of 5000; the dotted curve is for a larger Reynolds Number or for a stream of greater turbulence at a Reynolds Number of 5000.

E. Location of separation of laminar boundary layer and general character of flow at Reynolds Number of 80000. Transition occurs simultaneously with separation.

F. Location of separation of turbulent boundary layer and general character of flow at Reynolds Number of 1000000. Transition occurs in the laminar boundary layer before separation.

irregular random motion superposed on the mean flow of the fluid by honeycombs, grids, or other upstream obstacles. The fluctuations of speed are of fairly high frequency and amount to only a few per cent of the mean speed at the most. A turbulence of only a few tenths of a per cent of the mean speed often has a

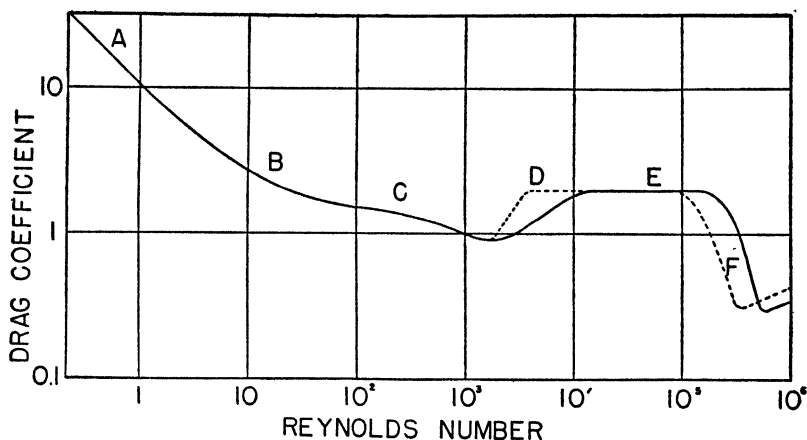


FIG. 2. Variation of drag coefficient of infinitely long circular cylinder with the Reynolds Number.

pronounced effect on the flow pattern. The curves shown correspond to intensities of turbulence of the order of 1 and 3 per cent.

Fig 1A shows the general character of the flow at extremely low Reynolds Numbers, i.e., less than 1. The inertia forces are small in comparison to the viscous forces in the region near the cylinder, and the flow closes in smoothly behind the cylinder. The flow resembles superficially the potential flow of a perfect fluid. The stream lines, however, for the same flux are at a greater distance from the cylinder than in the case of potential flow. The frictional effects extend to a distance of many times the diameter of the cylinder (2).

The drag coefficient is very large but decreases rapidly with increasing Reynolds Number. The force itself varies more nearly as the first power of the speed than as the second.

As the Reynolds Number is increased, the inertia forces become comparable with and finally larger than the viscous forces, and the character of flow gradually changes. Fig. 1B shows the flow at a Reynolds Number of 20 as computed by Thom (3) from the

Navier-Stokes equations of motion of a viscous fluid. Photographs by Thom, by Nisi, and by Fage (4) of the flow of air around a cylinder at Reynolds Numbers from 8 to 170 confirm the presence of the two stationary symmetrical eddies which become unstable as the Reynolds Number is increased, first vibrating irregularly, then breaking away alternately from the two sides. In the neighborhood of the point S occurs the first of the characteristic phenomena—that called separation. The main flow separates from the surface of the cylinder. Close to the cylinder the direction of the flow reverses on passing the neighborhood of S. At these Reynolds Numbers the rate of decrease of the drag coefficient becomes less rapid.

At a Reynolds Number only a little above 20, the exact value apparently depending on various factors such as the initial turbulence and the distance from the cylinder to other objects, the eddies begin to break away alternately on either side in a periodic manner to form the well-known Kármán vortex trail (5). Fig. 1C shows one stage of the flow obtained from one frame of a motion picture taken at a Reynolds Number of 174 by Nayler and Frazer (6). It would require many sketches to show the complete flow since the flow varies with the time. At this Reynolds Number the effect of viscosity is largely confined to a region near the cylinder about one-tenth of the diameter of the cylinder in thickness. This region is called the boundary layer. The speed decreases to zero at the surface of the cylinder. The thickness of the boundary layer decreases with increasing Reynolds Number.

If a traverse is made with a pitot tube behind the cylinder as indicated by AA' in Fig. 1C, the reference pressure being the pitot pressure in the approaching fluid stream, results are obtained (7) as shown in the auxiliary figure accompanying Fig. 1C. The falling total pressure in the wake of the cylinder is associated with a decrease in the speed of flow. The rapid changes are confined to a limited region whose width may be defined as indicated. The width of this mixing region, which is a continuation of the boundary layer beyond the separation point and accordingly designated a "free" layer, increases slowly with distance from its point of origin at the separation point S. While the flow is not steady with time, the flow in this free layer as well as in the attached layer is essentially of the type described by the Navier-Stokes equations in that the lateral mixing is due solely to the action of viscosity, the unsteadiness being in the nature of a wavering of the layer as



a whole. A type of motion in which lateral mixing is due to viscous action is usually termed laminar, and hence both the free layer and the attached layer are said to be laminar layers.

At Reynolds Numbers from 5,000 to 15,000 a new phenomenon was observed by Schiller and Linke (7). The thickness of the free layer after increasing slowly with distance from the separation point for some distance begins to increase at a more rapid rate, and as the Reynolds Number is increased the place at which the rapid increase in thickness begins moves closer and closer to the cylinder. Fig. 1D shows the boundaries of the free layer for Reynolds Numbers of 5,000 and 14,480 as determined by Schiller and Linke. We have learned to associate this rapid increase in thickness with transition from laminar to turbulent flow, i.e., from a type of motion in which the lateral transfer of momentum is by viscous shear to a type in which the lateral transfer occurs by the irregular motion of small portions of fluid. The change in character of flow in the wake modifies the pressure distribution in such a way that the drag coefficient increases with increasing Reynolds Number.

Schiller and Linke made the observation that the transition occurs closer to the cylinder at a given Reynolds Number when the turbulence of the stream was increased by inserting a wire screen of small mesh. Under the same condition the rise in the drag coefficient occurred at a lower Reynolds Number. This sensitiveness to the effect of turbulence is also a well-known characteristic of the transition from laminar to turbulent flow.

At some value of the Reynolds Number between 5,000 and 20,000, the exact value being dependent on the initial turbulence, the point of transition in the free layer reaches the laminar separation point and for a considerable range of Reynolds Number transition begins at the separation point. The drag coefficient is constant throughout this range. The location of the separation point and the general nature of the flow are indicated schematically in Fig. 1E. Periodic fluctuations of speed of the type usually associated with a Kármán vortex trail are found in the wake. However, no Kármán trail has been seen or photographed at Reynolds Numbers above 2,500, i.e., above those at which the trail results from the rolling up of a laminar layer. When turbulence sets in, the vorticity is rapidly diffused as is any substance used as an indicator in photographing the flow. Vortices, however,

break away periodically for Reynolds Numbers up to 100,000 or more.

When the Reynolds Number reaches 100,000 or more, the value again being dependent on the initial turbulence, transition begins in the boundary layer ahead of the laminar separation point. Separation is then delayed because of the more intense mixing which sweeps away a large part of the wake. Separation of the turbulent boundary layer still occurs, though delayed. The periodicity in the wake disappears and the whole wake becomes turbulent, the motions being random and highly irregular. The drag coefficient decreases to a much lower value. The flow is indicated schematically in Fig. 1F.

## CHARACTERISTIC PHENOMENA

The important characteristic phenomena which are noted in the flow about a cylinder and which can be found in many other flow patterns may be listed as follows:

1. *Formation of free and boundary layers*, i.e., regions in which intense changes in velocity are concentrated, the velocity gradients being much larger than elsewhere. At high Reynolds Numbers boundary layers are found on the upstream portions of the surfaces of immersed solid bodies and often at other boundary surfaces. A free layer may occur elsewhere in the flow, usually as a continuation of a boundary layer attached to a surface. It can be recognized by the presence of velocity gradients much greater than elsewhere in the flow.

2. *Laminar flow*, i.e., a flow in which the lateral transfer of momentum occurs solely through the action of viscosity. It can be recognized by the small rate of diffusion of material particles, streams of smoke or dye maintaining their identity for relatively long distances, by the relatively slow obliteration of velocity gradients, and the relative thinness of the free layers and boundary layers. The average flow conforms to that computed from the Navier-Stokes equations.

3. *Turbulent flow*, i.e., a flow in which the lateral transfer of momentum occurs by the irregular motion of small portions of fluid. It can be recognized by the presence of irregular and random velocity fluctuations of relatively high frequency, by the rapid rate of diffusion of material particles or heat, and by the relatively thick free layers and boundary layers.

4. *Transition*, i.e., the change from laminar to turbulent flow occurring in some limited part of the field of flow. It can be recognized by the change in the rate of thickening of free or attached layers, by a marked increase in the rate of diffusion of material particles and heat, by a marked increase in frequency and loss of periodic characteristics in the velocity fluctuations, and by an acceleration of the flow close to the boundaries of immersed objects.

5. *Separation*, i.e., a separation of the main flow from a boundary accompanied by a reversal in direction of the flow very close to the boundary behind the separation point and by the formation of a wake in which the velocity is much reduced. It is recognized most easily by the reversal which can be detected by emitting suitable tracer substances from openings in the boundary.

It should be noted that transition may occur either in a free layer or in an attached boundary layer, but that it occurs more readily in free layers. So far as is known, turbulence in streams of constant destiny arises only in free or attached layers.

Likewise, separation may occur either with a laminar or a turbulent layer, but occurs more readily in laminar layers.

## THE STUDY OF TRANSITION

The most common occurrence of transition in the experience of the average person is that in the rising column of smoke from a cigarette lying on an ash-tray in a quiet room. The smoke rises for some distance in smooth filaments which may wave around but do not lose their identity. This flow is a laminar flow. At some distance above the cigarette, the filaments suddenly break up into a confused eddying motion, a turbulent motion. The transition moves closer to the cigarette when the air of the room is disturbed.

The simplest flow in which transition has been observed and studied quantitatively is that in the boundary layer of a thin smooth flat plate set parallel to the fluid flow, the so-called skin-friction plate (8). It cannot yet be said even in this simple case that transition is fully understood, although much progress has been made. Techniques have been developed not only to detect the occurrence of transition by the rapid thickening of the layer and by the increase of the velocity close to the plate, but also to measure the distribution of the mean velocity and the intensity of the velocity fluctuations. The experimental data at hand (9)

show that in this case the principal parameters which determine the location of transition are the speed and viscosity of the fluid and the intensity and scale of the initial turbulence.

If the plate is placed in a stream in which the static pressure increases or decreases, the location of transition depends on the sign and magnitude of the pressure gradient. A pressure decreasing downstream is favorable to the maintenance of laminar flow, while a pressure increasing downstream accelerates transition.

Experiments under other conditions indicate that curvature of the surface and surface roughness also influence transition, concave or rough surfaces causing transition further upstream, while convex surfaces show transition further downstream than plane surfaces.

The attempt to generalize the quantitative information now available has not been fully successful. Transition does not seem to depend solely on the local condition of the boundary layer as described by its thickness, the speed of the fluid stream just outside it, the intensity and scale of the turbulence just outside, the local pressure gradient, and the curvature and roughness of the surface. The subject is under active study and it is hoped that at least the simplest case of transition will soon be understood.

As an illustration of a more detailed study of transition and the recognition of the several characteristic phenomena, Fig. 3 summarizes the results of a survey of the speed distribution in the boundary layer of a cylinder of elliptic cross section with major and minor axes 11.78 and 3.98 inches, respectively, subjected to an air flow parallel to the major axis at a mean speed of 70 ft/sec with an initial turbulence amounting to 0.85 per cent of the mean speed. The curves in the figure are contour lines of equal mean speed derived from measurements with a hot-wire anemometer which is insensitive to the direction of air flow.

The interpretation and a full discussion of these measurements are given in Technical Report No. 652 of the National Advisory Committee for Aeronautics. The boundary layer is laminar or nearly so to a distance of about 10.8 inches from the nose. Separation of the laminar boundary layer occurs at about 10 inches from the nose as compared to a value of 9.5 inches computed from the theory of laminar boundary layers. Transition of the free layer is evidenced by the rapid approach of the contours toward the surface beginning at a distance of 10.8 inches from the nose. In this region the speed near the surface increases on proceeding down-

stream. The layer in fact reattaches itself to the surface and separates again as a turbulent layer at 12.1 inches from the nose.

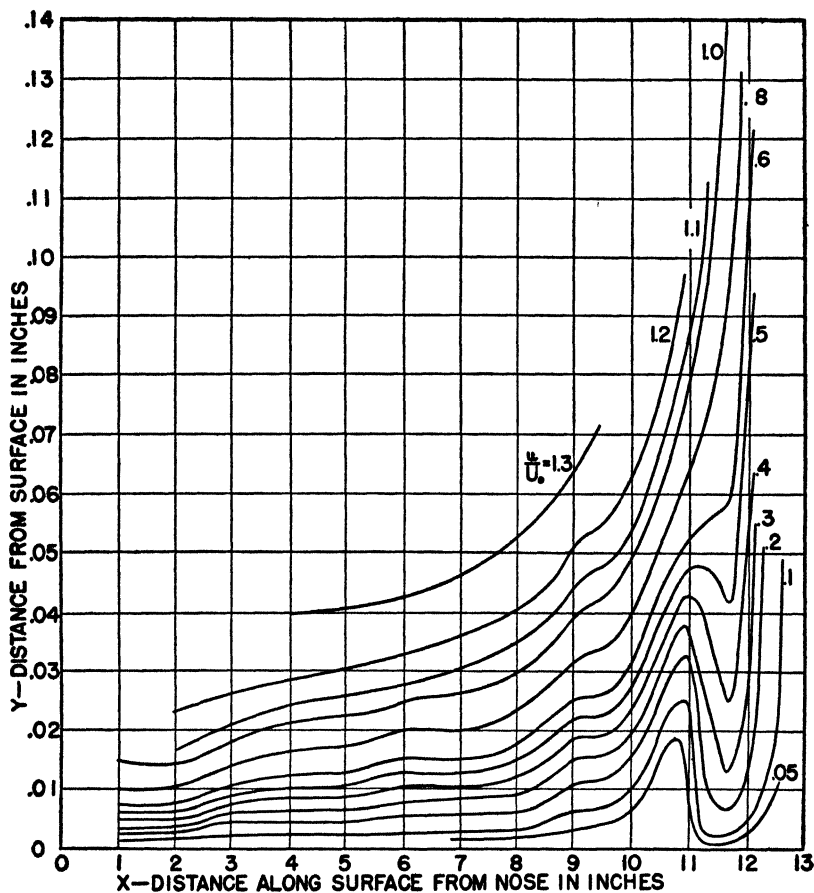


FIG. 3. Distribution of mean speed in the boundary layer of a 3.98 by 11.78 inch elliptical cylinder in an airstream speed 70 ft/sec, intensity of turbulence 0.85 percent, scale of turbulence 0.26 inch.

$U_0$  is the speed at a large distance from the cylinder,  $u$  is the local speed. The curves are contours of constant speed.

Laminar separation occurs at  $x = 10$ , transition of the free laminar layer at 10.8 followed by reattachment of the layer and separation of the turbulent layer at 12.1.

## RÔLE OF TRANSITION

In general terms transition may be stated to play both beneficial and detrimental rôles in fluid mechanics. It is beneficial in that it

delays separation of the flow and promotes more rapid diffusion of momentum, matter, and heat. It is detrimental in that it increases skin-friction and often modifies the pressure distribution in such a way that drag is increased. These general effects may be observed in practically every field of application of fluid mechanics.

In aeronautics, for example, the lift of the wings of an airplane is very much reduced and their drag greatly increased if separation of the flow occurs on the upper surface. The occurrence of transition is beneficial in that it causes the air flow to close in around the wings delaying separation until close to the trailing edge. The lift of the wings is maintained to relatively high angles of attack and the drag is much less than would be the case with separation of a laminar boundary layer. The desired performance of wings, struts, and fuselages could not be obtained if separation were not delayed by the transition in the boundary layer from laminar to turbulent flow. On the other hand, transition is detrimental in that it increases the skin-friction of wings, fuselages, and propeller blades. High-speed performance is greatly affected by the location of transition on these members. In aeronautical engineering it is desirable therefore to be able to control the position of transition so that it is delayed as long as possible to avoid increase in the frictional resistance but that it occurs early enough to avoid flow separation.

Another important rôle of transition in aeronautics is in its relation to the interpretation of model experiments. Small-scale turbulence is absent in the atmosphere but is present in varying degree in wind tunnels. Because of the effect of turbulence on transition great care must be exercised in the conduct and interpretation of wind tunnel tests.

In marine, automotive, and railroad engineering transition plays the same rôle in determining the resistance to forward motion as in aeronautics. Towing basins as well as wind tunnels have varying degrees of initial turbulence, and it is the practice in many towing basins to control the position of transition on ship models by adding local surface roughness in the form of sand grains or wire bands at the desired transition point.

In chemical engineering transition plays an important part in problems of the flow of fluids in pipes or other enclosed channels, in problems of diffusion and mixing, and in problems of heat transfer. The relationships are very different in laminar and turbu-

lent flow. The occurrence of transition increases pressure losses in flow in pipes, increases the rate of diffusion of matter and the rate of transfer of heat.

In hydraulic engineering, transition plays a part in the transport of sand and sediment, the scouring of river beds, the separation of the flow around obstacles, around the blades of water turbines, and in draft tubes and other diffusers, friction in pipes and channels, and various mixing processes such as those involved in the disappearance of density currents.

Any of these applications could be described in some detail, although in many cases the concepts outlined herein have not been applied to guide the experimental work. Special techniques are required to identify the occurrence of the characteristic phenomena, but it is by no means necessary to make tedious velocity traverses very close to the surface as illustrated by Fig. 3. Such traverses supply a great deal of additional information but simple methods are sufficient for mere identification of separation and transition.

### REFERENCES

1. RELF, E. F. Discussion of the results of measurements of the resistance of wires, etc. Brit. Adv. Comm. Aeronautics, Repts. and Memo. No. 102; Annual Reports for 1913-14, 5:47, 1915.  
WIESELSBERGER, C. Weitere Feststellungen über die Gesetze des Flüssigkeit- und Luftwiderstandes. Physik. Z., 23:219, 1922.
2. FAGE, A., AND J. H. WARSAP. The effects of turbulence and surface roughness on the drag of a circular cylinder. Brit. Aero. Res. Comm., Repts. and Memo. No. 1283, 1930.
3. SCHILLER, L., AND W. LINKE. Druck- und Reibungswiderstand des Zylinders bei Reynoldsen Zahlen 5000 bis 40000. Z. Flugtechnik u. Motorluftschiffahrt, 24:193, 1933.
4. PRANDTL, L. The mechanics of viscous fluids. 11. Theory of very slow motions. Div. G of W. F. Durand's Aerodynamic Theory. (Springer, 1935) Vol. III, p. 69-80.
5. THOM, A. The flow past circular cylinders at low speeds. Proc. Roy. Soc. (London), 141:651, 1933.
6. NISI, H., AND A. W. PORTER. On eddies in air. Phil. Mag. (6), 46:754, 1923.  
NISI, H. Experimental studies on eddies in air. Japanese Jour. Physics, 4:1, 1925.
7. FAGE, A. Photographs of fluid flow revealed with an ultra-microscope. Proc. Roy. Soc. (London), 144:381, 1934.

5. VON KÁRMÁN, TH., AND H. RUBACH. Über die Mechanismus des Flüssigkeit- und Luftwiderstandes. Physik. Z., 13:49, 1912.
6. NAYLER, J. L., AND R. A. FRAZER. Vortex motion, etc. Brit. Adv. Comm. Aeronautics, Repts. and Memo. No. 332, 1917.
7. SCHILLER, L., AND W. LINKE. Druck- und Reibungswiderstand des Zylinders bei Reynoldsen Zahlen 5000 bis 40000. Z. Flugtechnik u. Motorluftschiffahrt, 24:193, 1933.
8. DRYDEN, H. L. Air flow in the boundary layer near a plate. Nat. Adv. Comm. Aeronautics, Tech. Rept. No. 562, 1936.
9. DRYDEN, H. L. Turbulence and the boundary layer. Jour. Aero. Sci., 6:85, 1939.





UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

## Problems of Flow in Compressible Fluids

By , .

THEODORE VON KÁRMÁN, PH.D., DR. ING., SC.D.\*

IN MANY applications of fluid mechanics the assumption of incompressibility or constant density of the fluid yields results of sufficient accuracy for practical purposes. In general it is assumed that this approximation is justified if the velocity, density, and pressure changes are relatively small, as, for example, in most problems of practical hydraulics and aeronautics. Often it is stated that the dynamics of incompressible fluids will give satisfactory results if the velocities involved in the problem are small in comparison with the velocity of sound. However, if we look over the entire field of fluid mechanics somewhat more carefully, we find that besides the high-speed phenomena there are other cases in which density variation or elasticity of the fluid cannot be neglected. In all, the problems in which compressibility enters as a governing factor can be classified in the following groups:

a) *Pressure propagation.* In an incompressible fluid the pressure changes propagate with infinite velocity. Hence, if we are concerned with pressure oscillations in fluids, we have to consider the elasticity of the fluid as the factor governing the phenomena, irrespectively of the magnitude of the velocities or the pressure changes involved. Problems of acoustics, water hammer in pipes, oscillations in gas conduits, engine manifolds, etc., belong in this class of problems.

b) *Density currents.* In the problems belonging in this group we are concerned with a stratified fluid medium, as the atmosphere or water carrying silt. In such problems the velocities and pressure changes caused by the motion may be small, but the phenomena are governed by the gravity field produced by density differences.

c) *Slow motion of fluids with large density changes.* In this class of problems the pressure and density changes may be produced by some mechanism, for example, by the pistons of a reciprocating

\* Director of the Daniel Guggenheim Aeronautical Laboratory and Professor of Aeronautics, California Institute of Technology.

engine. In other cases the change of density is caused by large pressure gradients, as, for example, in the flow of gases through granular media.

d) *High-speed flow and high-speed motion of solids in fluid media.* In flow problems belonging in this class the density and pressure differences are produced or accompanied by large variations of the velocity. This class includes the flow of elastic fluids through ducts, vane systems, etc., with velocities of the order of or higher than the velocity of sound, and the motion of solid bodies with such velocities. Such problems occur in the design of gas and steam turbines, compressors, propellers, jet propulsors, high-speed airplanes, and in particular in exterior ballistics.

This paper is restricted to a short review of a few problems belonging to item d). They deserve a great deal of attention because a better understanding of the high-speed phenomena is essential for further progress in machinery dealing with elastic fluids, in high-speed aeronautics, and in shell and bomb design.

We shall use the term "gas dynamics" for this group of fluid mechanics problems. It is understood that gas dynamics is based on combination of the principles of fluid mechanics and thermodynamics.

## 1. GAS DYNAMICS IN "HYDRAULICS" FASHION

It has become fashionable in the last few years to make a distinction between hydraulics and hydrodynamics (or fluid mechanics). It is hard to say exactly where hydraulics ends and fluid mechanics begins. Some authors consider hydraulics as a purely empirical science and hydrodynamics as a purely mathematical discipline dealing with ideal fluids. I believe this is a rather erroneous view, since hydraulics is based on the same principles of mechanics which are the essential content of the equations of hydrodynamics, and hydrodynamics has never restricted its domain to ideal fluids. It seems to me that what we might call treatment of fluid mechanics problems in the "hydraulic manner" is something similar to the approximate treatment of problems of elastic bodies by the ordinary beam theory. In both cases field distributions, i.e., two- and three-dimensional velocity and stress distributions, are approximated by essentially one-dimensional patterns, using certain standard distributions and average values across the cross sections of a pipe, channel, or beam.

Let us consider, for example, the flow of a fluid through a nozzle of variable cross section. To fix the ideas we assume that the fluid is at rest in a container under certain pressure and flows into a space of zero absolute pressure. The mass flow through the nozzle is given by the product  $\rho Av$  where  $A$  is the cross section area,  $v$  the mean velocity of the fluid through the cross section, and  $\rho$  its density. Hence, the condition of continuity requires that  $\rho Av = \text{const.}$  In the case of an incompressible fluid the density is constant and therefore we have  $Av = \text{const.}$  or

$$\frac{dA}{A} = -\frac{dv}{v} \quad (1)$$

In the case of an elastic fluid

$$\frac{dA}{A} = -\frac{dv}{v} - \frac{d\rho}{\rho} \quad (2)$$

If we neglect friction losses, the dynamic equation for stationary flow can be written in the form ( $p$  = pressure)

$$v dv + \frac{dp}{\rho} = 0$$

or

$$\frac{d\rho}{\rho} = -\frac{1}{\frac{dp}{d\rho}} v dv \quad (3)$$

It is known that the quantity  $\frac{dp}{d\rho}$  is equal to the square of the velocity of propagation of infinitesimally small pressure difference, in other words it is equal to the square of the velocity of sound  $a$  in a gas of the given pressure  $p$  and density  $\rho$ . Hence

$$\frac{d\rho}{\rho} = -\frac{v}{a^2} dv \quad (4)$$

Substituting Eq. (4) in Eq. (2) we obtain

$$\frac{dA}{A} = -\frac{dv}{v} \left( 1 - \frac{v^2}{a^2} \right) \quad (5)$$

or introducing the ratio  $\frac{v}{a}$  as the "local Mach's number"  $M$ ,

$$\frac{dA}{A} = -\frac{dv}{v} \left( 1 - M^2 \right) \quad (6)$$

It is seen

- 1) that the approximation of incompressibility holds if  $M^2 \ll 1$ .
- 2) that the cross section corresponding to a stationary flow through the nozzle decreases with increasing velocity if  $M^2 < 1$  and increases with increasing velocity if  $M^2 > 1$ .
- 3) that at the cross section of minimum area, i.e., "at the throat,"  $M = 1$ ; the velocity is equal to the "local velocity of sound."

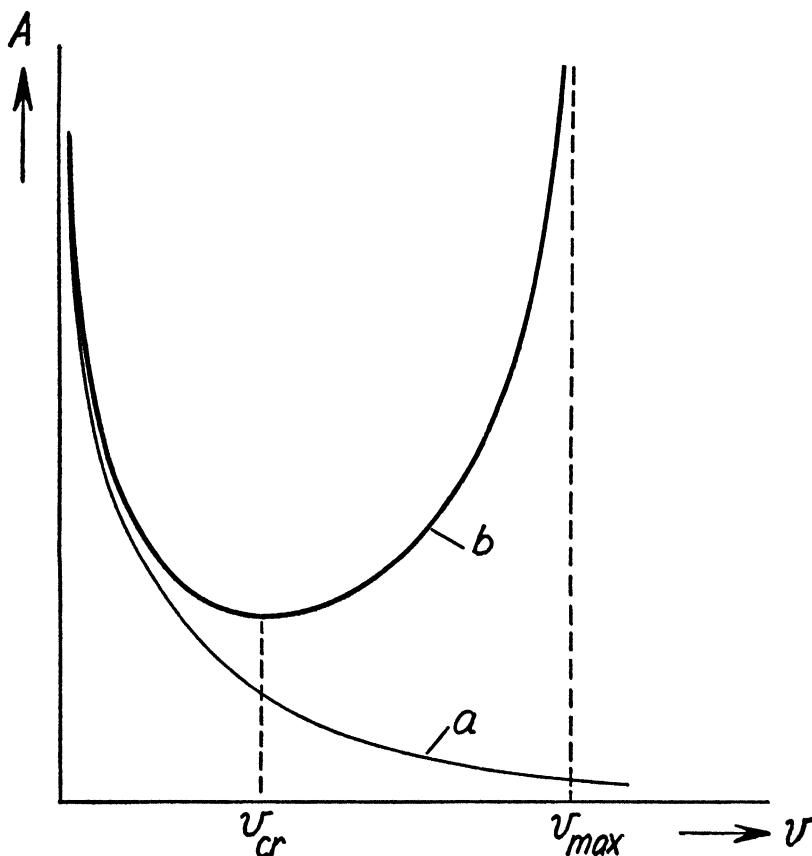


FIG. 1

Figure 1 shows the distribution of cross section area as function of the velocity  $v$  first for an incompressible fluid (line  $a$ ) then for an elastic fluid (line  $b$ ) under the assumption of adiabatic change.

In this case  $a^2 = a_0^2 - \frac{\gamma - 1}{2} v^2$  where  $\gamma = \frac{c_p}{c_v}$  is the ratio between the specific heats for constant pressure  $p$  and constant volume  $V$ ;  $a_0$  is the velocity of sound corresponding to the initial state at rest.

Let us now assume that a certain nozzle has been designed for a pressure ratio  $\frac{p_0}{p_*} = n > n_{cr}$  where  $n_{cr}$  is the ratio between initial pressure and the pressure at the throat. The question arises what happens if the pressure  $p_0$  is maintained at the entrance of the nozzle, but the pressure at the exit is varied between the limits  $p_0$  and  $p_*$ .

First, it is evident that the fluid, after reaching the velocity of sound at the throat, can either continue the isentropic expansion or can be isentropically compressed in the diverging portion of the nozzle. Evidently the flow capacity of the nozzle is the same in both cases; however, in the second case the gas leaves the exit with a velocity  $v_e' < a$ . This velocity and the pressure at the exit are the same as in the cross section of the same area in the converging portion of the nozzle. Denoting this pressure by  $p_e'$  we see that we obtain smooth flow with maximum capacity for two values of the pressure, namely  $p_*$  and  $p_e'$ . If the exit pressure is higher than  $p_e'$ , i.e., it is between the limits  $p_0$  and  $p_e'$ , we obtain subsonic isentropic flow whose capacity increases with the pressure difference between entrance and exit. If the exit pressure is between  $p_e'$  and  $p_*$ , no isentropic flow is possible that is consistent with both the pressure difference and the exit cross section area. In such cases a more or less discontinuous change of pressure takes place somewhere between the throat and the exit or in the free jet behind the exit. Such discontinuous change is called a "shock wave" for reasons to be explained in the next section.

If we assume that the discontinuous pressure change takes place at a certain cross section and behind this section the process is isentropic again, we can compute isentropic flows with new values of the entropy so that they are consistent with the given end pressure and the exit cross section area. Figure 2 shows a family of pressure curves corresponding to such isentropic flows. Figure 3 shows the corresponding states of the fluid in the pressure volume diagram. If the discontinuous transition ends on the upper branch of one of these pressure curves, the gas is compressed and its flow is subsonic downstream of the shock; a jump ending on a lower branch would mean that the gas expands and flows with supersonic velocity. It will be seen later that the second case is physi-

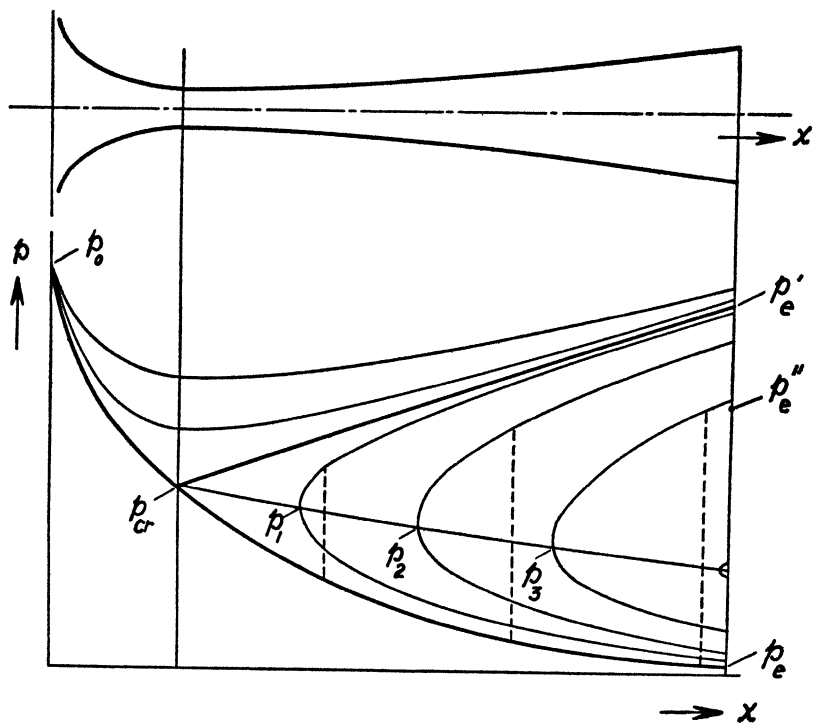


FIG. 2

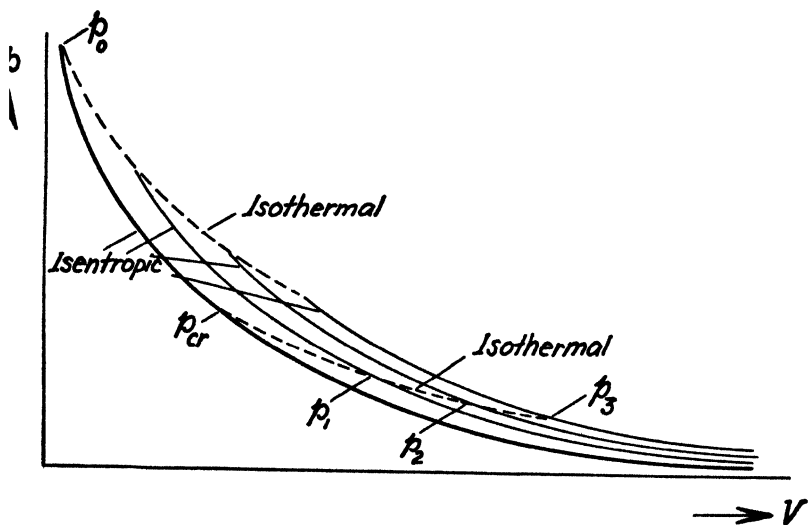


FIG. 3

cally not possible. In fact, there is a minimum value  $p_e''$  of the end pressure, which can be obtained by such a shock; this is the value obtained when the shock takes place at the exit. If the end pressure is smaller than  $p_e''$ , the shock occurs in the free jet in the form of so-called "oblique" shock waves. It can also occur that the stream separates from the wall. Which process takes place depends probably on the shape of the nozzle and frictional effects. If the end pressure is between  $p_e'$  and  $p_e''$ , a plane shock wave is observed in the diverging part of the nozzle. The magnitude and location of this discontinuity depends on the dynamical relations to be discussed below.

## 2. SHOCK WAVES

The velocity of sound in a fluid is the velocity of propagation of infinitesimally small pressure changes. It has been mentioned that the square of the velocity of sound is equal to the derivative of the pressure  $p$  with respect to the density  $\rho$ . In an ideal case under the assumption of isentropic change  $\frac{dp}{d\rho} = \gamma \frac{p}{\rho}$ , where  $\gamma$  is the ratio between the specific heats  $c_p$  and  $c_v$ , Earnshaw and Riemann investigated the propagation of finite disturbances. In particular Riemann emphasized that in a compressible fluid finite discontinuities of pressure can be developed by a process which itself is continuous.

Let us consider one dimensional wave motion in an elastic fluid originally at rest. The dynamic equation can be written in the form

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{1}{\rho} \frac{\partial p}{\partial x} = 0 \quad (2.1)$$

or

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + \frac{dp}{d\rho} \frac{1}{\rho} \frac{\partial \rho}{\partial x} = 0 \quad (2.2)$$

The equation of continuity is given by

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0 \quad (2.3)$$

If we consider small perturbations, i.e., we assume that  $\rho' = \rho - \rho_0$



( $\rho_0$  = density at rest) and  $u$  are small quantities, and put  $\frac{d\rho}{d\rho} = a^2$ , we obtain by neglecting terms of higher order:

$$\begin{aligned}\frac{\partial u}{\partial t} &= -\frac{a^2}{\rho_0} \frac{\partial \rho'}{\partial x} \\ \frac{\partial \rho'}{\partial t} &= -\rho_0 \frac{\partial u}{\partial x}\end{aligned}\quad (2.4)$$

It is easily seen that for example,

$$\begin{aligned}u &= u_0 \phi(x - at) \\ \rho' &= \rho_0 \frac{u_0}{a} \phi(x - at)\end{aligned}\quad (2.5)$$

where  $\phi$  is an arbitrary function, is a solution of this system of differential equations. The equations (2.5) represent a plane wave of arbitrary shape proceeding in the positive  $x$  direction with the velocity of propagation  $a$ .

It can be easily verified that the wave motion represented by equations (2.5) satisfies the equation:

$$\frac{\partial u}{\partial t} = -a \frac{\partial u}{\partial x} \quad (2.6)$$

and that  $\frac{\rho'}{\rho_0} = \frac{u}{a}$ . Now if we assume that also in the case of a finite disturbance the variation of the density is a function of the velocity only, it can be shown that equation (2.6) will be replaced by

$$\frac{\partial u}{\partial t} = -(a + u) \frac{\partial u}{\partial x} \quad (2.7)$$

Consequently, if the distribution of  $u(x)$  is given by the first equation (2.5) at  $t = 0$ , for example, the shape of the velocity distribution curve will be deformed for  $t > 0$ . Whereas in the case of infinitesimally small waves the velocity distribution curve is displaced unchanged by a length  $a\Delta t$  in the time  $\Delta t$ , in the case of the finite wave the displacements of the positive ordinates are larger and those of the negative ordinates are smaller than  $a\Delta t$ .

Hence, the velocity gradient  $\frac{\partial u}{\partial x}$  and the density gradient  $\frac{\partial \rho}{\partial x}$  increase during the propagation of the wave. If we apply this consideration, for example to a wave of sinusoidal shape [i.e., if  $\phi(x - at)$  is a sine function], we find that the tendency to develop

an abrupt change in density occurs at places where the fluid moves in the direction of increasing density and pressure. After a certain time  $\frac{\partial \rho}{\partial x}$  becomes infinite and thereafter the equations fail to give a continuous solution for the velocity and density distributions.

The process is somewhat analogous to the collapse of ocean waves breaking on the shore due to the fact that the velocity of propagation of the wave crest is greater than that of the trough.

Assuming now that pressure, density, and velocity undergo a discontinuous change at a plane  $x = \text{constant}$ , the velocity of propagation of the discontinuity can be computed by satisfying the following conditions: (a) conservation of matter through the discontinuity surface, (b) conservation of momentum, and (c) conservation of energy. These relations are established in the easiest way if the discontinuity surface is considered stationary and the fluid passing through it. In other words we consider the discontinuity surface as the front of a standing wave known as a shock wave. In this case pressures and densities on the two sides of the shock wave satisfy the following relation:

$$\frac{\Delta p}{\Delta \rho} = \frac{p_2 - p_1}{\rho_2 - \rho_1} = \gamma \frac{p_2 + p_1}{\rho_2 + \rho_1} \quad (2.8)$$

(The subscript 1 refers to the state before, 2 behind the shock wave.)

It is seen that if  $\rho_1 \rightarrow \rho_2 \rightarrow \rho$  and  $p_1 \rightarrow p_2 \rightarrow p$ , this relation becomes

$$\frac{dp}{d\rho} = \gamma \frac{p}{\rho} \quad (2.9)$$

This expression is the differential equation for isentropic change. However, if  $\Delta p$  and  $\Delta \rho$  are finite, the entropy of the gas is larger on the side of the shock wave where the pressure is larger. Therefore according to the second law of thermodynamics the gas passes through the shock wave from the low-pressure to the high-pressure side. In other words the progressing wave with finite discontinuity is always a compression wave: it propagates from regions of high pressure to regions of low pressure. It can also be shown that the velocity of propagation of a finite discontinuity relative to the fluid on the low-pressure side is always greater than the velocity of sound in that medium. Therefore, a standing shock wave can exist only if the fluid upstream of the shock wave moves with

supersonic velocity. The velocity of propagation relative to the high pressure side is smaller than the velocity of sound in that medium. In other words the fluid passing through a standing shock wave changes its condition from supersonic to subsonic. Thus it is clear why no plane shock wave can occur between the initial pressure curve and one of the lower branches in the nozzle considered in the preceding section.

For a real fluid no mathematical discontinuity can exist. The thickness of the transition layer has been estimated by several authors, for example, by L. Prandtl, Lord Rayleigh, R. Becker, and G. I. Taylor. It appears that it is determined chiefly by the viscosity and the heat conduction of the gas. The calculations of the authors mentioned give for large compression ratios extremely small values for the thickness, values of the order of  $10^{-5} - 10^{-7}$  cm, so that it is doubtful whether in this case the hydrodynamic equations and the equations of heat conduction can be applied inside of the transition layer. For pressure ratios near to unity, i.e., for the case that the change of velocity,  $\Delta u$ , of the gas passing through the shock wave is small in comparison to the velocity with which the gas enters the shock wave, the thickness of the transition layer can be expressed, according to G. I. Taylor, by the simple relation  $\delta = \text{const} \frac{\nu}{\Delta u}$  where the constant is of the order 6.5 and  $\nu$  is the kinematic viscosity of the gas and it is assumed that the ratio  $\frac{c_p \mu}{\lambda}$  ( $\mu$  = coefficient of viscosity,  $c_p$  = heat capacity,  $\lambda$  = coefficient of heat conduction) is constant and is slightly lower than unity. For example, for air of atmospheric pressure and for a velocity change of  $\Delta u = 30$  m/sec,  $\delta = 0.003$  mm.

These calculations are based on the assumption that the statistical equilibrium of the gas is reached instantaneously, i.e., we can assume a statistical equilibrium no matter how fast the change of state takes place. Now the time in which the gas passes through the transition layer is, for example, in the case considered above, of the order of  $10^{-8}$  sec if the velocity of the gas is of the order of the velocity of sound. For larger compression ratios and velocities the time will be of the order of  $10^{-10}$  sec. This time interval appears to be too small to secure statistical equilibrium. In particular the energy distribution of the intermolecular degrees of freedom will have considerable time lag, which is, for example, of the order of

$10^{-6}$  sec for nitrogen. It appears that an adequate analysis of the structure of the shock wave requires a treatment from the point of modern statistical mechanics.

### 3. THE ANALOGY BETWEEN OPEN CHANNEL FLOW AND FLOW OF ELASTIC FLUIDS

The phenomenon of the shock wave is to some extent analogous to the hydraulic jump observed in open channel flows. The hydraulic jump is an abrupt transition between so-called supercritical and subcritical flow. The flow of water in an open channel is supercritical if the mean velocity  $v$  is larger than the velocity of long waves in a shallow channel, which is given by  $\sqrt{gh}$  ( $g$  = acceleration of gravity,  $h$  = depth of the flow). The flow is called subcritical if the mean velocity is under this critical value. The continuity equation for stationary flow has the form

$$vh = \text{constant} \quad (3.1)$$

This corresponds to the continuity equation of an elastic gas through a pipe of constant cross section:

$$v\rho = \text{constant} \quad (3.2)$$

if we let the density  $\rho$  of the gas correspond to the depth  $h$  of the flow.

The dynamic equation for channel flow is

$$gdh + vdv = 0 \quad (3.3)$$

The dynamic equation for a gas:

$$\frac{dp}{d\rho} \frac{d\rho}{\rho} + vdv = 0 \quad (3.4)$$

We remember that  $\frac{dp}{d\rho}$  is the square of the velocity of sound; hence

$\frac{dp}{d\rho}$  corresponds to  $gh$ , and replacing  $\rho$  by  $h$  we see that  $\frac{dp}{d\rho} \frac{d\rho}{\rho}$  will be replaced by  $gdh$ .

It is seen that both the continuity equation and the dynamic equation are analogous, if we consider  $h$  and  $\rho$  as corresponding quantities. The quantity  $P = \frac{gh^2}{2}$ , which is the resultant of the static pressure acting on a cross section of the fluid, corresponds to the pressure  $p$  of the gas.

The analogy between the one-dimensional flow of an elastic fluid and open-channel flow of an incompressible fluid can be extended to the two-dimensional case. We find that the equations of the strictly two-dimensional flow of an elastic fluid and the approximate equations for the motion of a layer of an incompressible heavy fluid of variable thickness are identical if we replace the density of the gas  $\rho$  by the thickness  $h$  of the incompressible fluid layer and the pressure of the gas by the quantity  $P = \frac{gh^2}{2}$ .

However, assuming adiabatic change, the relation between the pressure and density of the gas has the form  $P = \text{const.} \rho^\gamma$ , whereas for the fluid layer  $P = \text{const.} h^2$ . Hence, the flow of the fluid layer corresponds to the flow of a gas with  $\gamma = 2$ . Nevertheless, the analogy is quite useful for quantitative study of some complicated flow problems. This analogy was first given by E. Jouguet; it was developed by D. Riabouchinsky and E. Preiswerk. The present writer applied the same analogy to illustrate wave phenomena in curved channels.

#### 4. PROBLEMS OF SUBSONIC FLOW IN AERONAUTICS

One of the most surprising developments in modern fluid mechanics is the successful application of the theory of potential motion of ideal incompressible fluids to the actual flow of air around airfoils and streamlined bodies. A few decades ago the potential theory was considered as a purely mathematical discipline. Later it was understood that it might be a valuable guide for understanding of the general laws which govern the lift, moment, induced drag, etc., of airfoils. However, as aerodynamical design became more refined, it was found that the quantitative predictions of the theory are of a remarkable accuracy in spite of its idealized conception. As a matter of fact, the aeronautical designer has today a great confidence in the results of the theory of potential motion. It is possible to design airfoil sections and wings for certain given pressure and lift distributions. One notable limitation is the phenomenon of stalling, which essentially depends on the velocity distribution in the immediate neighborhood of the solid surface, in other words, on the behavior of the boundary layer.

However, when the velocity of an aircraft reaches about one-half of the velocity of sound, discrepancies appear due to the

neglection of the compressibility of the air, and it becomes imperative to extend the theory to elastic fluids.

Let us consider two-dimensional flow. In this case the motion of the fluid is completely defined by the stream function and the velocity potential function or, in other words, by the streamlines and the lines of equal potential. These two families of curves are orthogonal, and in the case of an incompressible fluid they constitute a pattern of geometrically similar square-surface elements. The mathematical expression of this condition can be written in the form:

$$\frac{\partial \phi}{\partial x} = \frac{\partial \psi}{\partial y}, \quad \frac{\partial \phi}{\partial y} = -\frac{\partial \psi}{\partial x} \quad (4.1)$$

where  $\phi(x, y)$  is the potential function and  $\psi(x, y)$  is the stream function. The corresponding equations for an elastic fluid have the form

$$\frac{\rho}{\rho_0} \frac{\partial \phi}{\partial x} = \frac{\partial \psi}{\partial y}, \quad \frac{\rho}{\rho_0} \frac{\partial \phi}{\partial y} = -\frac{\partial \psi}{\partial x} \quad (4.2)$$

where  $\rho$  is the variable density of the fluid  $\rho_0$  is its density at rest. The density is a function of the absolute value of the velocity; it is given by the relation

$$\int_{\rho_0}^{\rho} \frac{d\rho}{\rho} + \frac{1}{2} \left\{ \left( \frac{\partial \phi}{\partial x} \right)^2 + \left( \frac{\partial \phi}{\partial y} \right)^2 \right\} = \text{const.} \quad (4.3)$$

$\frac{d\rho}{d\rho}$  is equal to the square of the velocity of sound and is a given function of the density  $\rho$ .

The solution of the system of equations (4.2) and (4.3) is infinitely more complicated than the solution of the system (4.1). The latter leads immediately to Laplace's equation both for  $\phi$  and  $\psi$ , and we can use the powerful and elegant methods of conformal transformation. In the case of (4.2) and (4.3) the differential equation for  $\phi$  has the form

$$\left(1 - \frac{u^2}{a^2}\right) \frac{\partial^2 \phi}{\partial x^2} - \frac{2uv}{a^2} \frac{\partial^2 \phi}{\partial x \partial y} + \left(1 - \frac{v^2}{a^2}\right) \frac{\partial^2 \phi}{\partial y^2} = 0 \quad (4.4)$$

where  $u = \frac{\partial \phi}{\partial x}$ ,  $v = \frac{\partial \phi}{\partial y}$  and  $a^2 = a_0^2 - \frac{\gamma - 1}{2} (u^2 + v^2)$ . The quantity  $a_0^2$  is a constant, equal to the square of the velocity of sound

in the gas at rest. The complicated nonlinear equation (4.4) cannot be expected to allow elegant and easy methods of solution.

The equation for  $\phi$  becomes linear if we consider  $u$  and  $v$ , i.e., the velocity components, as independent variables. In other words we draw equipotential lines  $\phi = \text{const.}$  in the "hodograph plane." The equation for  $\phi$  as function of the velocity components resulting from this transformation is a linear one, and it is possible to find particular solutions for it. For example, if we introduce polar co-ordinates in the hodograph plane, putting  $u = q \cos \theta$ ,  $v = q \sin \theta$ , where  $q$  is the absolute magnitude of the velocity and  $\theta$  is the inclination between the velocity vector and the  $x$ -axis, the differential equation for  $\phi$  has particular solutions of the form  $\phi = f(q)e^{in\theta}$  where the function  $f(q)$  can be expressed by hypergeometrical functions and can be calculated with sufficient accuracy by means of convergent power series. However, to obtain the corresponding flow pattern and velocity distribution we have to transfer the streamlines into the physical  $x, y$  plane. Then we sometimes shall find that solutions which look beautiful in the hodograph plane give, for example, overlapping flows in the physical plane. In other words our method is an inverse one, and an assumed relationship between the stream function and the velocity components does not necessarily give a flow with physically possible boundaries. However, in a few cases the method has yielded useful results. A further promising application of the hodograph method to approximate solutions of flow problems is discussed later in this section.

Turning our attention to approximate methods of solution, the first idea that occurs is to apply the process of iteration. We might start from equation (4.1) putting  $\rho = \text{constant}$ , say  $\rho_0$  in (4.2). Then we calculate  $\rho(x, y)$  from equation (4.3) and substitute this function in equation (4.2). This general idea was applied in different forms by Lord Rayleigh, O. Janzen, S. G. Hooker, L. Poggi, C. Kaplan, and I. Imai analytically, and by G. I. Taylor experimentally. The last-mentioned author used an electric computing device for the solution of (4.2). This device consists essentially of a layer of an electrolyte of variable thickness, the variable thickness being determined by the function  $\rho(x, y)$ . The analytical iteration process employed by the other authors requires relatively much labor and it is difficult to have an estimate of the degree of approximation achieved by the consecutive steps.

A rather simple and ingenious method of approximation is

based on the assumption that the flow of a fluid around an airfoil or a streamlined body can be considered as a small perturbation of a parallel flow. If we denote the velocity and density of the undisturbed flow by  $U$  and  $\rho^0$ , and assume that  $u$  and  $\rho$  are only slightly different from  $U$  and  $\rho^0$  respectively and  $v$  is small in comparison to  $U$ , we can write  $\phi = Ux + \phi'$ ,  $\psi = \rho^0 Uy + \psi'$  and neglect terms containing squares and products of  $\phi'$ ,  $\psi'$ ,  $\rho'$  and their derivatives.

In this case the equations (4.2) become

$$\frac{\rho^0}{\rho^0} \frac{\partial \phi'}{\partial x} + \frac{\rho'}{\rho^0} U = \frac{\partial \psi'}{\partial y}, \quad \frac{\rho^0}{\rho^0} \frac{\partial \phi'}{\partial y} = - \frac{\partial \psi'}{\partial x} \quad (4.6)$$

and equation (4.3) takes the form:

$$a^{02} \frac{\rho'}{\rho^0} + U \frac{\partial \phi'}{\partial x} = 0 \quad (4.7)$$

Substituting (4.7) in the first equation (4.6) we obtain

$$\left(1 - \frac{U^2}{a^{02}}\right) \frac{\partial \phi'}{\partial x} = \frac{\partial \psi'}{\partial y}$$

$$\frac{\partial \phi'}{\partial y} = - \frac{\partial \psi'}{\partial x} \quad (4.8)$$

Eliminating  $\psi'$  from these two equations we have

$$\left(1 - \frac{U^2}{a^{02}}\right) \frac{\partial^2 \phi'}{\partial x^2} + \frac{\partial^2 \phi'}{\partial y^2} = 0 \quad (4.9)$$

Equation (4.9) can be applied in the following way. Suppose we have solved the problem of the flow of an incompressible fluid around a cylindrical body of a certain cross section. Then the same velocity potential gives the flow of a compressible fluid around a cylindrical body whose cross section is affine to the cross section of the first body and can be derived from it by expansion in the flow direction in the ratio  $1: \sqrt{1 - M^2}$ , where  $M^2 = \frac{U^2}{a^{02}}$  and  $U$  is the velocity of the undisturbed fluid. This consideration was suggested by H. Glauert and L. Prandtl.

In the case of a streamlined body or an airfoil section the designer is mainly interested in finding the magnitude and the location of the highest velocity. It is assumed in general that if the velocity at any place reaches the velocity of sound (more exactly



the local velocity of sound), shock waves appear, the lift of the airfoil decreases, and its drag increases very considerably. If the thickness of the airfoil section and the camber of its center line are small in comparison to the chord length, and the section is contracted in the ratio  $\sqrt{1 - M^2}$ : 1 in the length direction, the maximum value of the suction which determines the maximum velocity increases in the ratio  $\frac{1}{\sqrt{1 - M^2}}$ . Hence, if the maximum suction

produced in an incompressible fluid is equal to  $\beta \rho^0 \frac{U^2}{2}$ , we obtain

in the case of a compressible fluid a suction equal to  $\frac{\beta}{\sqrt{1 - M^2}} \rho^0 \frac{U^2}{2}$ .

Denoting the velocity increase by  $u$ , we have approximately the relation  $\rho^0 U u = \frac{\beta}{\sqrt{1 - M^2}} \rho^0 \frac{U^2}{2}$  or

$$u = \frac{\beta}{2} \frac{U}{\sqrt{1 - M^2}} \quad (4.10)$$

The velocity will reach the velocity of sound, if  $U + u = a^0$ , or  $\frac{u}{a^0} = 1 - M$ . Hence, the critical Mach's number, i.e., the ratio  $\frac{U}{a^0}$  for which the local velocity reaches the velocity of sound is given by the relation

$$1 - M = \frac{\beta}{2} \frac{M}{\sqrt{1 - M^2}} \quad (4.11)$$

or

$$(1 - M)^{3/2} (1 + M) = \frac{\beta M}{2} \quad (4.12)$$

For very small value of  $\beta$ ,  $1 - M$  is small and we have

$$M = 1 - \left( \frac{\beta}{4} \right)^{2/3} \quad (4.13)$$

The comparison with the experimental evidence shows that equation (4.12) underestimates the effect of compressibility; the equation gives too large values for the critical Mach's number. C. Kaplan suggested using a more exact relation between the velocity increase and suction. Then, he believes, the discrepancy is eliminated; however, it appears from the experiments that this correc-

tion is not sufficient. Moreover, from the theoretical point of view it is rather unsafe to mix up in the same consideration a linear approximation and relations which take higher terms into account.

The present writer believes that closer and more consistent approximation can be obtained by an adequate adaptation of a method used by A. Chaplygin, B. Demtchenko, A. Busemann and others for the flow of compressible fluids at low velocities. This adaptation has been suggested by the writer and worked out in an elegant way by H. S. Tsien. The method makes use of the transformation discussed above as the hodograph method. We consider the potential function and the stream function as functions of the velocity components. For sake of convenience we use the absolute magnitude,  $q$  and the inclination  $\theta$  of the velocity vector (i.e., the angle between the direction of the undisturbed flow and the velocity vector) as polar co-ordinates in the hodograph plane. Then the equations for  $\phi$  and  $\psi$  in the case of an incompressible fluid have the form

$$q \frac{\partial \phi}{\partial q} = - \frac{\partial \psi}{\partial \theta}, \quad \frac{\partial \phi}{\partial \theta} = q \frac{\partial \psi}{\partial q} \quad (4.14)$$

In other words, the complex quantity  $\phi + i\psi$  is an analytical function of the complex variable  $\log q - i\theta$ . Let us now approximate the isentropic curve  $p = \text{const} \left(\frac{1}{\rho}\right)^{-\gamma}$  in the neighborhood of the point  $p^0, \rho^0$  by a linear approximation

$$p - p^0 = \gamma \left( \frac{1}{\rho^0} - \frac{1}{\rho} \right) p^0 \rho^0 \quad (4.15)$$

Then it can be shown that the equations for a compressible fluid whose physical behavior corresponds to equation (4.15) appear without further approximation in the form:

$$\frac{\rho^0}{\rho} q \frac{\partial \phi}{\partial q} = - \frac{\partial \psi}{\partial \theta}, \quad \frac{\partial \phi}{\partial \theta} = \frac{\rho^0}{\rho} q \frac{\partial \psi}{\partial q} \quad (4.16)$$

If we now introduce a new variable

$$\omega = \int_{\rho^0}^{\rho} \frac{\rho}{\rho^0} \frac{dq}{q} \quad (4.17)$$

which is a generalization of  $\log q$ , the variable used above, we have

$$\frac{\partial \phi}{\partial \omega} = -\frac{\partial \psi}{\partial \theta}, \quad \frac{\partial \phi}{\partial \theta} = \frac{\partial \psi}{\partial \omega} \quad (4.18)$$

Hence,  $\phi + i\psi$  will be an analytical function of  $\omega - i\theta$ . In this case we are able to find solutions of the system of equation (4.18) using the method of conformal transformation as it is used in the case of incompressible fluids.

This method can be employed for the determination of the pressure distribution in a compressible fluid, provided the pressure distribution in an incompressible fluid is known. Let us assume that  $\varphi(\omega, \theta)$  and  $\psi(\omega, \theta)$  are the potential function and the stream function respectively for the flow of an incompressible fluid around a given cross section, where  $\omega$  is the logarithm of the magnitude and  $\theta$  the angle of inclination of the velocity. Now if the values of  $q$  corresponding to  $\omega$  according to equation (4.17) are taken as the values of the velocity, we obtain a flow pattern satisfying the equations (4.16) of the compressible fluid. The relation (4.17) enables us to establish a simple relation between the magnitudes of the suction produced in an incompressible and a compressible fluid; this relation depends on the Mach's number  $M = \frac{U}{a^0}$ . Denoting the maximum suction produced in an incompressible fluid by  $-\Delta p$ , we have for the suction produced in a compressible fluid

$$-\Delta \bar{p} = \frac{-\Delta p}{\sqrt{1 - M^2} + \frac{M^2}{2(1 + \sqrt{1 - M^2})}\Delta p} \quad (4.19)$$

It is seen that for very small values of  $\Delta p$  the result is identical with the result obtained by use of the Glauert-Prandtl method. However, for larger values of the suction the effect of Mach's number appears intensified by the influence of the second term in the denominator. As a matter of fact  $\Delta \bar{p} \rightarrow \infty$  for a value of  $M$  which is smaller than unity.

If we return to the physical plane and draw the flow pattern mentioned above, we notice that the section undergoes a slight deformation. This transformation has been carried out by Tsien for several cases, and it seems that the influence of this slight distortion of the cross section does not change the results to a significant extent. Moreover, the correction is likely to be balanced by

the deviation between the exact isentropic relation and the approximate relation (4.15) used in this theory. It would be of great interest to check the predictions of the theory sketched above with a great number of measurements. Insofar as measurements are available on elliptic cylinders and airfoils, good agreement has been found.

Flow problems in steam and gas turbines offer an as yet unexplored field for applications of the theory of flow of compressible fluids. Many design problems related to such machinery are concerned with the flow of a compressible fluid through a lattice. Hence they require an extension of the methods discussed above and applied heretofore to single airfoils to an infinite system of identical airfoils.

## 5. APPLICATIONS OF THE THEORY OF COMPRESSIBLE FLUIDS TO EXTERIOR BALLISTICS

The problem of the drag of projectiles is one of the oldest problems of fluid mechanics and has attracted the attention of a great number of mathematicians and physicists, beginning with Isaac Newton. Newton's theory of air resistance was later abandoned. As a matter of fact, the basic discrepancy between experience and d'Alembert's paradox, according to which a body moving with uniform velocity through a non-viscous fluid does not meet any resistance, to some extent blocked the way to a rational theory of drag. The various fantastic propositions for best shapes of projectiles reproduced in Figs. 4 and 5 show the utmost lack of any theoretical guide to empirical speculation.<sup>1</sup> To be sure the fundamental distinction between drag at supersonic and subsonic speeds has been pointed out by quite a number of physicists and ballisticians dealing with the problem; namely, that the energy loss corresponding to the drag at supersonic speeds is dissipated in the waves accompanying the projectile, in particular in the head wave emanating from the nose of the bullet, whereas the energy loss at subsonic velocities is mainly dissipated in eddies produced in the rear of the body. However, even an approximate quantitative theory of the wave resistance of shells was first attempted only a few years ago, and an exact theory does not exist as yet.

<sup>1</sup> From Th. H. Green, *Army Ordnance*, May-June 1932.

The calculation of the retardation of projectiles by air resistance requires the solution of the problem of axially symmetric flow of

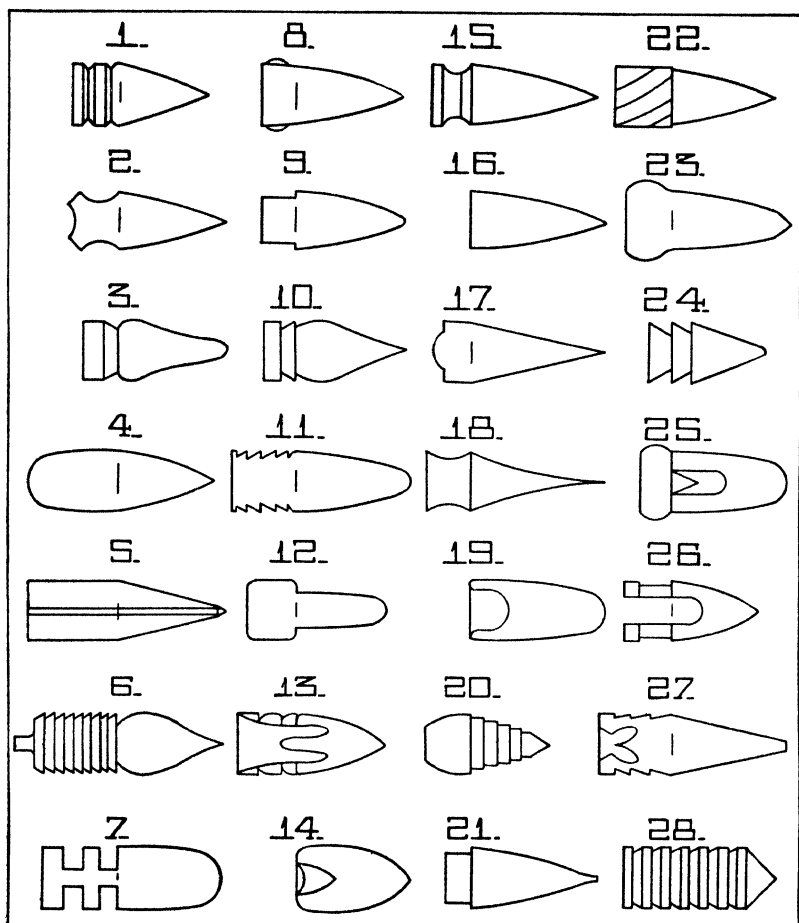


FIG. 4. Small arms projectile designs prior to the American Civil War.

1. Prussian; 2. Swiss; 3. Russian; 4. Prussian; 5. Arrow Ball; 6. Spanish; 7. Hunting; 8. Russian; 9. Colts; 10. Russian; 11. Mangeot; 12. Richards; 13. Peters; 14. Beckwith; 15. Russian; 16. American; 17. English; 18. Sardinia; 19. Enfield; 20. Pomme à Pin; 21. Mexican; 22. Whitworth; 23. Swiss; 24. Prussian; 25. Beckwith; 26. French; 27. Mangeot; 28. Tamisien.

an elastic fluid around a body of revolution for various velocities. It is not difficult to write up the differential equation for such a flow, as we did in the case of two-dimensional flow (cf. Eq. (4.4)).

However, the equation for the case of axial symmetry is even more complicated than the equation (4.4). Hence, we will consider first

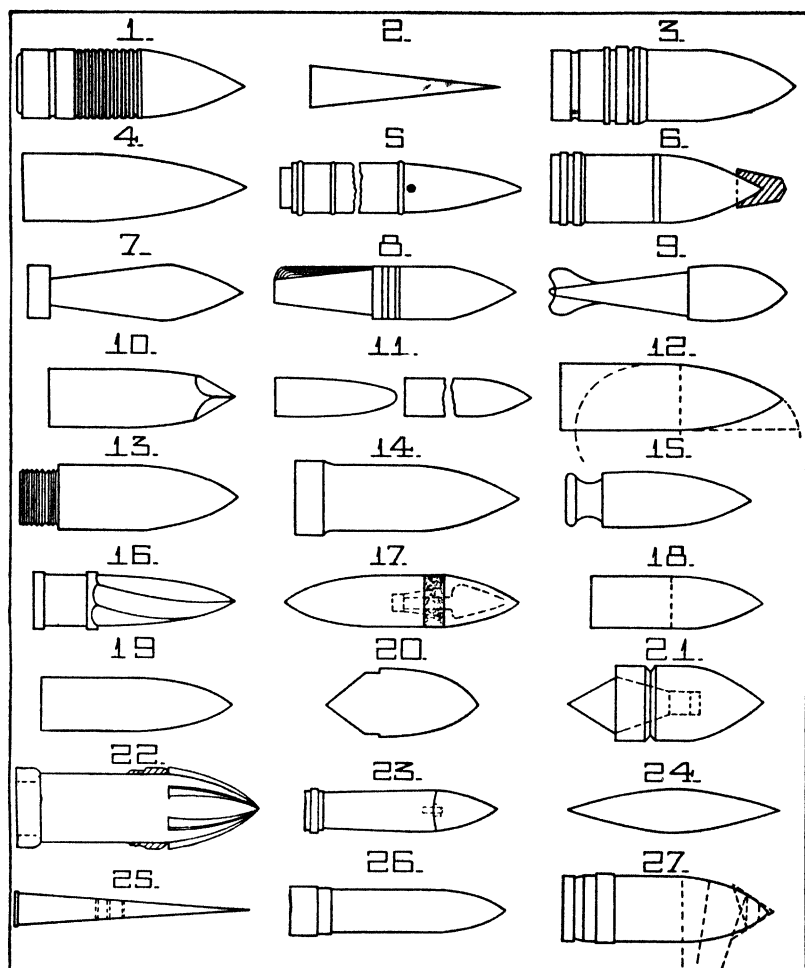


FIG. 5. The form of some of the projectiles patented in Great Britain.

1. Brandon; 2. Farini; 3. Grenfell; 4. Trebor; 5. Unge; 6. Reichwald; 7. Hinton; 8. Parry; 9. Shipright; 10. Cole; 11. Simpson; 12. Hadfield; 13. Edmonds; 14. Lindenufer; 15. Johnson; 16. Fisher; 17. Parsons; 18. Palliser; 19. Mennens; 20. Lake; 21. Hazeltine; 22. Palliser; 23. Longsdon; 24. Hope; 25. Farini; 26. Day; 27. Page.

the approximate linearized equation which corresponds to equation (4.9). Then we easily obtain the following differential equation

tion for the potential  $\phi$  as function of the cylindrical co-ordinates  $r$  and  $x$ :

$$\left(1 - \frac{U^2}{a^2}\right) \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} = 0 \quad (5.1)$$

If  $U < a^0$ , the solution of this equation can be reduced to the solution of Laplace's equation, as was explained in the previous section. However, if  $U > a^0$  as in most problems in exterior ballistics, the equation is of the type of the wave equation:

$$\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} \quad (5.2)$$

which controls the propagation of small disturbances in a two-dimensional medium.

In order to visualize the physical meaning of equation (5.2), let us assume that in a two-dimensional medium initially at rest a small "radial burst" of the character of a source at  $r = 0$  occurs at the time  $t = 0$ . The effect of this source will propagate in the radial direction so that at the time  $t$  the portion of the medium for which  $r > ct$  will be undisturbed and the effect of the disturbance will be restricted to the portion  $r < ct$ . Therefore, if we choose  $t$  and  $r$  as cylindrical co-ordinates, the effect of the disturbance applied at  $r = 0$  at the time  $t = 0$  will be restricted to the inner portion of a cone whose vertex is at  $r = t = 0$  and whose half vertex angle is equal to  $\tan^{-1} c$ . The surface of the cone contains the circles  $r = ct$  which form the wave front at various times. If we apply this conclusion to equation (5.1), we have to keep in

mind that  $c$  is replaced by  $\frac{1}{\sqrt{\frac{U^2}{a^2} - 1}}$ . Therefore, if we assume a

stationary disturbance at an arbitrary point on the  $x$ -axis, say  $r = 0$ ,  $x = x_0$ , the effect of such a disturbance will be restricted to the inner portion of a cone whose vertex is at  $r = 0$ ,  $x = x_0$  and whose half vertex angle is equal to

$$\beta = \tan^{-1} \frac{1}{\sqrt{\frac{U^2}{a^2} - 1}} = \sin^{-1} \frac{a^0}{U} \quad (5.3)$$

The angle  $\beta$  is called Mach's angle. The relation (5.3) can be easily visualized in the following way. Assume that the fluid is at

rest and the disturbance moves in the negative  $x$  direction with the velocity  $U$ . Then at any instant a spherical wave will be created by the disturbance. Evidently the common front of these spherical waves is a cone with the half vertex angle equal to  $\sin^{-1} \frac{a^0}{U}$ , since in the time  $t$  the disturbance, i.e., the vertex of the cone, is displaced by the length  $\bar{U}t$  and the radius of the spherical wave is equal to  $a^0t$ .

The solution of the wave equation of the type (5.2) has been worked out by T. Levi-Civita and H. Lamb. Their solution has been used by the present writer in a joint paper with N. Moore for the computation of the flow around a projectile in the following manner: We assume continuously distributed sources along the  $x$ -axis between the nose and the base of the projectile and determine the distribution of sources in such a way that the velocity normal to the surface of the shell vanishes, at least with the approximation used in this linear theory. It is quite remarkable that the values of the drag calculated from the pressure distribution resulting from this simplified theory is in fair accordance with the drag vs. velocity curve derived from firing experiments. To be sure the accordance is only good in the case of slender and sharp-nosed projectiles. In the case of shells with blunt noses no such accordance can be expected since in that case no justification exists for the assumption of small disturbances.

This linearized theory has been used by the writer for the computation of the best shape of an ogive; recently C. Ferrari made interesting contributions to the same problem, which first was treated by Newton. H. S. Tsien has extended the theory to the flow around projectiles whose axis is inclined at a small angle to the flow direction. He used a distribution of doublets along the axis of the shell, and computed the coefficients of the side force and pitching moment. These quantities have important bearing on problems in exterior ballistics.

If we give up the assumption that the flow can be approximated by small deviations from a uniform parallel flow, we obtain in general at least one discontinuity surface which has the character of a standing oblique shock wave. A standing oblique shock wave is a discontinuity surface which the fluid enters in a direction different from the normal. In this case only the normal component of the velocity suffers an abrupt change on passing through the shock wave; the tangential component remains unaltered. There-



fore, the streamlines show a break at the discontinuity surface. We already mentioned such oblique shock waves in connection with the nozzle problem. In fact, if the pressure at the exit of a nozzle is slightly higher than the calculated end pressure, the gas may continue its isentropic expansion to the end section; then at the edges of the end section the higher value of the external pressure creates oblique shock waves which are reflected at the free surface of the jet, until the surplus energy is dissipated by impact losses and friction.

The exact shape of the oblique shock wave produced at the sharp edge of an obstacle can be computed exactly in the two-dimensional case by the methods of characteristics. Such computation was, for example, carried out by L. Crocco for a sharp-nosed half-body. As a matter of fact in the case of two-dimensional flow the solution of the exact equation of an elastic fluid is much simpler for supersonic than for subsonic flow due to the fact that in the supersonic case the equation has real characteristics.

However, in the case of three-dimensional flow, even if axial symmetry can be assumed, the solution is very laborious and the only case in which it has been worked out in detail is that of a conical obstacle of infinite extension. This problem was first treated by A. Busemann, who noticed that the partial differential equation can be reduced to an ordinary differential equation. G. I. Taylor and J. W. Maccoll calculated numerically the solutions of this equation and discussed the various flow patterns in great detail. It appears that one obtains conical standing shock waves above a certain Mach's number, whose magnitude depends on the vertex angle of the cone. It also appears that if the half vertex angle exceeds  $57.6^\circ$  no conical shock wave can be found at any velocity. In the range in which conical shock waves exist, there are two solutions corresponding to two different values of the inclination of the shock waves. However, it seems that only one of the two shock waves is stable.

The observations of the flow around sharp-nosed shells confirm the conclusion of the theory that shock waves can be found at the tip only above a certain minimum value of Mach's number. For smaller Mach's number it appears from the observations that the shock wave is detached from the nose of the shell and lies ahead of the nose of the projectile. It seems that the flow between the shock wave and the shell is subsonic at least near the axis of symmetry. It has not been possible as yet to compute the shape of such a

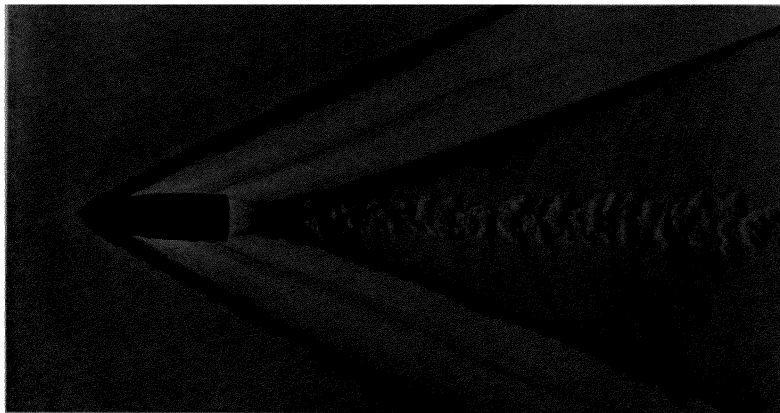


FIG. 6. Oblique shockwaves starting from the nose.

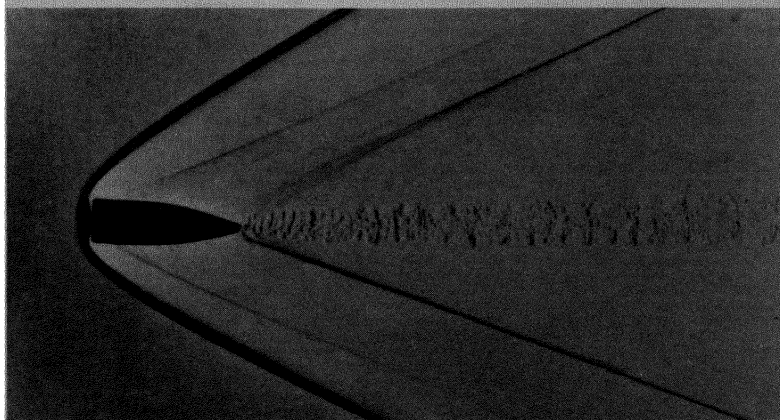


FIG. 7. Projectile fired backwards. Shockwave detached from the shell.



detached wave surface and the velocity and pressure distribution at the shell surface resulting from such flow conditions. Ferrari applied the method of characteristics to integration of the exact differential equation of the flow with axial symmetry around a shell for the case in which the flow is supersonic everywhere. This method requires cumbersome calculations, but leads in principle to the solution of the problem. For the mixed supersonic-subsonic case, which occurs at lower Mach's number, no proper mathematical method has even been indicated.

Optical observations show behind a projectile a system of shock and expansion waves and some eddy formations of periodic character (Figs. 6 and 7); there are only qualitative descriptive analyses available for this complex flow pattern.<sup>2</sup>

It appears that although the simplified mathematical analysis is a valuable guide for the solution of problems of high-speed aerodynamics related to exterior ballistics, the difficulties of an exact mathematical analysis and the complex character of the physical phenomena make further experimental research imperative. It is the firm conviction of the present writer that as in aeronautics proper the wind tunnel—in this case the supersonic wind tunnel—will establish itself as a most valuable tool of research. It is bound to supplement the results of the firing tests and contribute new suggestions for further improvements in design.

<sup>2</sup> Figs. 6 and 7 are reproduced from Crane's textbook on Ballistics. Fig. 6 shows oblique shockwaves starting from the nose Fig. 7 (projectile fired backwards) shows a shockwave detached from the shell.



UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

# Investigations of Liquid Turbulence and Suspended Material Transportation

*By*

ANTON A. KALINSKE, M.S.\*

THE movement by a fluid of solid particles heavier than the fluid itself is an important phenomena in many engineering problems. The transportation of dust and sand by wind and air currents occurs in problems with which the meteorologist and the mechanical engineer are concerned. The movement of sediment by flowing water has been and will be more so in the future, one of the major problems of the hydraulic engineer. No doubt there are also many chemical engineering problems where the transportation of various suspensions in liquids is of significance.

The mechanics of flowing fluids must of course be first understood and studied if we are to know more about the action of the fluid on foreign particles. The interaction between the moving fluid and the heavier foreign particles is quite complex and has led to the formulation of a great variety of empirical rules and formulas. Considering the transportation of sand and silt by flowing water, experimental observations indicate that it is possible to describe three different methods by which this movement takes place. In order to provide a general background for the specific problem with which this paper is concerned, a brief description of these various methods or phases of sediment transport will be given.

As water flows over a layer of unconsolidated sand grains a sufficient force may be exerted on the individual grains so as to displace them and cause them to roll along the channel bed. Such movement is referred to as "bed-load" transportation. Depending on the shape of the particles and the water velocity, some of them may become raised off the bed due to various hydrodynamic forces set up as the water flows over a particle. Once lifted off the bed

\* Assistant Professor and Research Engineer, Iowa Institute of Hydraulic Research, State University of Iowa.

the particle will travel with the water for a short distance and then drop down. On striking the bed it may cause another particle to be dislodged off the bed and raised a short distance into the main stream. This particular process is referred to as "saltation," and though little is known about the process quantitatively, it appears to be always present when sediment transport takes place. It was mentioned by Gilbert (1) in the report of his investigations, and has been vividly described by Bagnold (2) in regard to the transport of sand in deserts by high velocity winds.

Of course, as soon as a particle of sand is raised off the bed it immediately tends to fall back due to the action of gravity. However, this immediate dropping back to the bed may not occur if the fall velocity is counterbalanced by an upward component equal or greater in magnitude. Such upward velocity components normal to the mean direction of flow occur in turbulent flow. If the particles are kept in the main body of the liquid due to the turbulent velocity components, then the material transported in this fashion is called "suspended-load." In most rivers and canals in the midwestern and western states the major portion of the sediment is transported in this manner.

Since material is kept in suspension due to the turbulence of the fluid, it is logical to suppose that any analyses or studies of suspended material transportation will depend on our understanding of the turbulence mechanism. Within recent years our knowledge of this mechanism has increased tremendously, and because of this fact we have been able to carry forward our analytical and experimental studies of suspended material transportation at a rapid pace.

## A. GENERAL THEORY OF DIFFUSION IN A TURBULENT FLUID

The simplest concept of how settleable particles are kept in suspension is to visualize the mass of fluid as filled with numerous eddies of various sizes whirling about in an apparently unpredictable and random fashion. These eddies are capable of transferring heat, mass, momentum, foreign particles, etc., from one region of the fluid to another. The assumption has been made that the diffusion process due to the turbulence eddies is similar in general to molecular diffusion; the individual eddies performing the function of the molecules. Experience seems to indicate that there is a

close similarity between the two phenomena if the scale of the turbulent diffusion process is large compared to the size of the eddies.

The simplest case of diffusion which it is possible to study experimentally is that of the spread from a line or point source of heat, or of matter of the same specific gravity as the turbulent water. The general equation of diffusion in a single direction for a steady state is:

$$n = -D \frac{dN}{dy}, \quad (1)$$

where  $n$  is the amount of matter transferred across a unit area in unit time,  $N$  is the concentration of the matter per unit volume and  $D$  is a "diffusion coefficient." For the case of molecular diffusion it has been shown that:

$$D = \frac{1}{2} \frac{\bar{y}^2}{t}, \quad (2)$$

where  $\bar{y}^2$  is the mean square displacement of the individual particles (such as Brownian particles) after a time  $t$ .

If particles having the same specific gravity as the fluid are injected into a turbulent stream of the fluid moving at a uniform velocity, their spread can be definitely calculated if the laws of molecular diffusion apply. If the concentration at the source is  $M$  then the concentration at a point a distance  $x$  downstream and a distance  $y$  normal to the horizontal line passing through the point of injection is given by:

$$N = \frac{M}{\sigma_y \sqrt{2\pi}} e^{-y^2/2\sigma_y^2}, \quad (3)$$

where  $\sigma_y$  is  $\sqrt{\bar{y}^2}$  or the root mean square of the transverse travel of the particles. The value of  $\sigma_y$  depends on the downstream distance,  $x$ . However, for molecular diffusion, as indicated in Eq. 2,  $\bar{y}^2$  is equal to  $2Dt$ , where  $t$  can in this case be taken as  $x/U$ , where  $U$  is the mean velocity of the fluid stream. In other words the value of  $\bar{y}^2$  varies directly as the downstream distance.

Observations of the spread of small immiscible droplets injected into a turbulent water stream indicate that the value of  $\bar{y}^2$  does not vary directly as the downstream distance. In Fig. 1 is shown the actual variation. Note that the variation of  $\bar{y}^2$  for small values



of  $x$  is parabolic and gradually becomes linear for larger values of  $x$ . In other words  $\frac{dy^2}{dx}$  tends to attain a constant value.

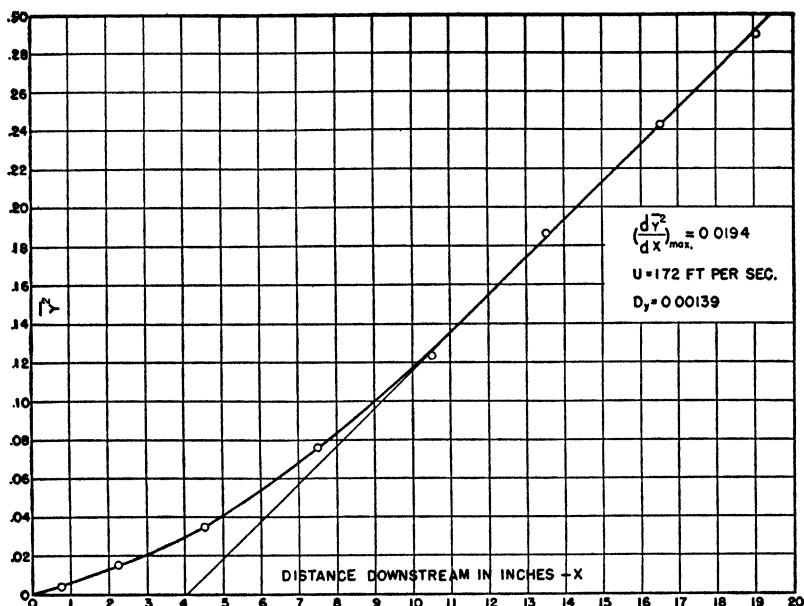


FIG. 1. Experimental data on diffusion in a turbulent water stream.

In order to explain analytically this phenomena Taylor (3) introduced the concept of a correlation coefficient,  $R_t$ , between the velocity of a fluid particle at any instant and its velocity at a time interval,  $t$ , later. With this concept Taylor develops the following relation for the variation of  $\bar{y}^2$  with  $t$ :

$$1/2 \frac{d\bar{y}^2}{dt} = \bar{v}^2 \int_0^t R_t dt, \quad (4)$$

where  $\bar{v}^2$  is the mean square value of the velocity component normal to the direction of flow. Thus the diffusion process is in essence reduced to a study of the variation of  $R_t$  with time. It appears plausible to suppose that in true turbulence  $R_t$  will approach zero as  $t$  increases, and when  $R_t$  is zero, the value of  $\int_0^t R_t dt$  becomes constant. This means that the variation of  $\bar{y}^2$  with  $t$  (or with  $x$  for a uniform mean velocity,  $U$ ) becomes linear.

A diffusion coefficient is defined for turbulence thus:

$$D = 1/2 \left( \frac{d\bar{y}^2}{dt} \right)_{\max.} = \frac{U}{2} \left( \frac{d\bar{y}^2}{dx} \right)_{\max.} \quad (5)$$

By obtaining motion pictures of the spread of small immiscible droplets injected at a point into turbulent water it is possible to calculate values of  $\bar{y}^2$  for various distances  $x$ , and thus determine the turbulent diffusion coefficient directly. Fig. 2 shows typical data of the variation of  $D$  from top to bottom in the center vertical section of an open water channel. The channel was 2.5 ft. wide and 1.0 ft. deep, and the mean velocity in the vertical section was .865 ft. per sec.

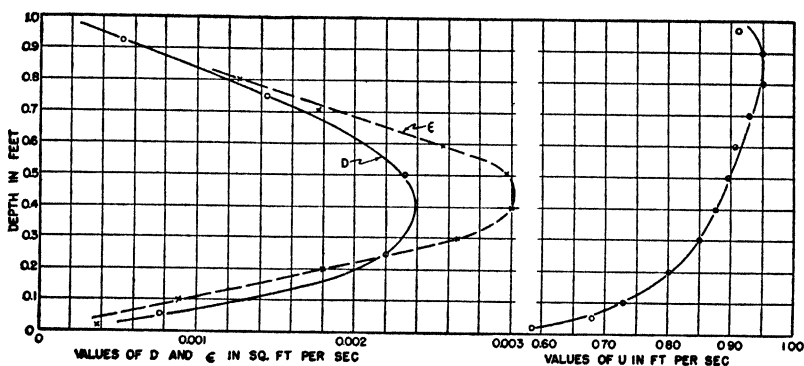


FIG. 2. Variation of mass and momentum diffusion coefficients and of mean velocity in the center of a smooth rectangular open channel.

The theory of momentum exchange due to the eddies in turbulent flow indicates that the shear in fully developed turbulence is:

$$\tau = \epsilon \rho \frac{dU}{dy} \quad (6)$$

where  $\epsilon$  is the momentum exchange coefficient and  $\rho$  the unit density. Whether the coefficient  $D$  is equal to  $\epsilon$  has not been definitely proven as yet, however, there is every reason to suppose that the two coefficients are at least proportional. In Fig. 2 are shown calculated values of  $\epsilon$  for the conditions for which  $D$  was determined experimentally. The computation of  $\epsilon$  was not very accurate since some assumptions had to be made regarding the value of the shear. The proximity of the numerical values of  $D$

and  $\epsilon$  indicates that these two coefficients are probably identical. More data is being obtained to check this further under various conditions.

## B. THEORY OF SUSPENDED MATERIAL TRANSPORTATION IN AN OPEN WATER CHANNEL

The analysis of the diffusion of sediment in turbulent water is complicated by the fact that it is constantly settling. If settleable material is placed in a turbulent liquid at a point, its diffusion will be quite similar to that of non-settling material, (Eq. 3), except that the point of maximum concentration instead of remaining at the same height is shifted down a distance equal to  $\frac{\pi c}{U}$ , where  $c$  is the velocity of fall of the sediment in still water.

The general differential equation controlling the phenomena of settleable material transportation in the main body of a turbulent fluid, considering just the two-dimensional case, is:

$$\frac{\partial N}{\partial t} = D_y \frac{\partial^2 N}{\partial y^2} + D_x \frac{\partial^2 N}{\partial x^2} + c \frac{\partial N}{\partial y} - \frac{U \partial N}{\partial x}, \quad (6)$$

where  $N$  is the mean concentration of sediment at any point  $(x, y)$  and at any time,  $t$ . The terms  $D_y$  and  $D_x$  are the diffusion coefficients in the designated directions. For the case when the sediment concentration in a channel does not vary on an average with time, and does not change downstream along the same elevation, Eq. 6 reduces to:

$$c \frac{dN}{dy} = D_y \frac{d^2 N}{dy^2}, \quad (7)$$

or,

$$cN = D_y \frac{dN}{dy} \quad (8)$$

Eq. 8 states that the net diffusion of sediment across a unit area in the  $y$ -direction by the turbulence is equal to the amount settling due to gravity. This is the condition of equilibrium; that is, the suspended sediment concentration does not vary in the direction of flow. In other words there is no filling or scouring of the channel bottom.

Integration of Eq. 8 gives:

$$\log_e \frac{N}{N_a} = -c \int_a^y \frac{dy}{D_y}, \quad (9)$$

where  $N_a$  is the concentration a distance  $a$  above the bottom. In order to evaluate  $D_y$  it is necessary to relate it to known hydraulic factors of open channel flow. For turbulent uniform flow with a mean velocity gradient normal to the direction of flow the unit shear is represented as:  $\tau = \epsilon \rho \frac{dU}{dy}$ , where  $\epsilon$  is the coefficient of momentum transfer. The assumption is made that  $\epsilon$  is equal to  $D_y$ . Then knowing the shear and the velocity gradient the value of  $D_y$  can be calculated. Mathematically this can be done by expressing the shear and the mean velocity as functions of the distance  $y$  from the bottom. For the case of a wide channel,  $\tau = \tau_o (1 - y/d)$ , where  $\tau_o$  is the bottom shear and is equal to  $\gamma dS$ , where  $\gamma$  is the specific weight of the fluid,  $S$  the channel slope and  $d$  its depth. The Prandtl-von Kármán law for velocity distribution in rough conduits, as transformed for open channels, is:

$$U/U_m = 1 + \frac{\sqrt{\tau_o/\rho}}{kU_m} (1 + \log_e y/d) \quad (10)$$

$U_m$  = mean velocity in vertical section,

$k$  = von Kármán's unusual constant, having a value between .38 and .40.

The value of  $\frac{dU}{dy}$ , using Eq. 10, is  $\sqrt{\tau_o/\rho}/ky$ . On substitution in the equation for shear the value of the exchange coefficient is obtained:

$$\epsilon = kd\sqrt{\tau_o/\rho}(1 - z)z, \quad (11)$$

where,  $z = y/d$ .

Note that  $\epsilon$  is zero at the bottom and the surface, and is a maximum at the mid-depth. Obviously,  $\epsilon$  does not become zero at the surface of most rivers, particularly those that show a definite "boiling." Also,  $\epsilon$  is not zero at the zero level of suspended material, since at this level the suspended material merges with the bed-load and no definite solid surface exists.

Substituting Eq. 11 for  $D_v$  in Eq. 9 and performing the integration we obtain:

$$\frac{N}{N_a} = \left( \frac{d/y - 1}{d/a - 1} \right)^{1/k} \quad (12)$$

$$t = c / \sqrt{\tau_o / \rho}$$

Though there is practically no experimental data for checking this relationship, Rouse (4) has presented some which appear to follow this relation quite closely. From a practical engineering standpoint Eq. 12 is not very convenient to use since the concentration is calculated as zero at the surface and infinity at the bottom. River measurements do not indicate that the suspended sediment concentration approaches zero at the surface. A more practically useful relation, and one that seems to approximate actual conditions quite well, is obtained if for the value of the diffusion coefficient,  $D_v$ , in Eq. 9, the average value of the coefficient in the vertical section is calculated. This average value can be calculated from Eq. 11:

$$\epsilon_{\text{ave.}} = d\sqrt{\tau_o / \rho} / 15 \quad (13)$$

(The factor 15 is obtained if  $k$  is assumed to be 0.40).

The equation for distribution of suspended material then becomes:

$$N = N_a e^{-15t(z-a)} \quad (14)$$

This relation plots as a straight line on semi-logarithmic paper, and the concentration at any point in the vertical section can be determined by simply plotting the known concentration,  $N_a$ , and drawing a straight line through this point at a slope as determined from  $-15t$ . In Fig. 3 are some actual measurements of suspended material concentrations in the Mississippi River at Muscatine, Iowa, as obtained with the assistance of the Rock Island U. S. Army Engineer Office. The straight lines drawn through the points indicating the concentration of sediment in a certain size range have a slope as calculated from Eq. 14, where  $t$  is equal to  $c / \sqrt{\tau_o / \rho}$ ,  $\tau_o$  is taken as  $\gamma dS$ , and  $c$  the average velocity of fall of the sediment in any particular size range. Other field data for the Missouri River and smaller rivers also seem to check the variation of suspended material as given by Eq. 14 very well.

In order to obtain the total amount of suspended sediment transported in unit time per unit width of a section of channel it

is necessary to multiply the concentration per unit volume at any point by the mean velocity at that point. This involves combining Eq. 14 with the velocity distribution equation (Eq. 10). The total

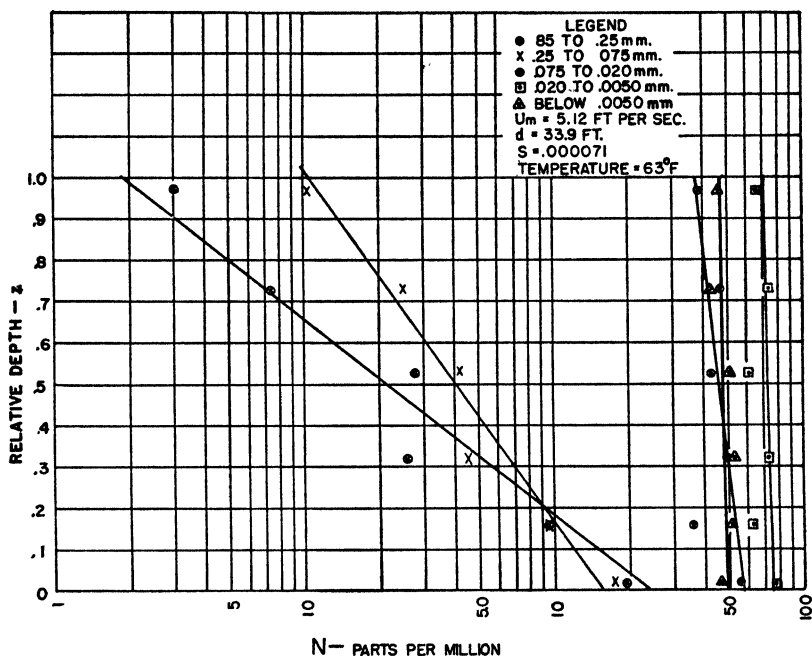


FIG. 3. Distribution of suspended material in a vertical section of the Mississippi River.

amount transported per unit time per unit width of any size range, characterized by the fall velocity,  $c$ , is then equal to:

$$\Sigma N_c = U_m d N_a e^{15ta} \phi(t, n/d^{1/6}) \quad (15)$$

The relative roughness ratio  $n/d^{1/6}$  comes into the function,  $\phi$ , since it can be shown that in Eq. 10 the quantity  $\sqrt{\tau_o/\rho}/kU_m = 1.70\sqrt{g} n/d^{1/6}$  when using Manning's formula for the mean velocity, and 0.40 for the value of  $k$ . The function,  $\phi(t, n/d^{1/6})$  which we will call  $P$ , is plotted in Fig. 4.

For the data shown in Fig. 3 the total amount of sediment carried of each size range was computed by multiplying the concentration at each point by the measured velocity, and was also computed using the bottom concentration as shown in Fig. 3 and then applying Eq. 15. The actual sum total of all sizes transported was

found to be 1.85 lbs. per second per unit width. The work involved in obtaining this figure was quite laborious since it involved a graphical integration for each size range. The value calculated by

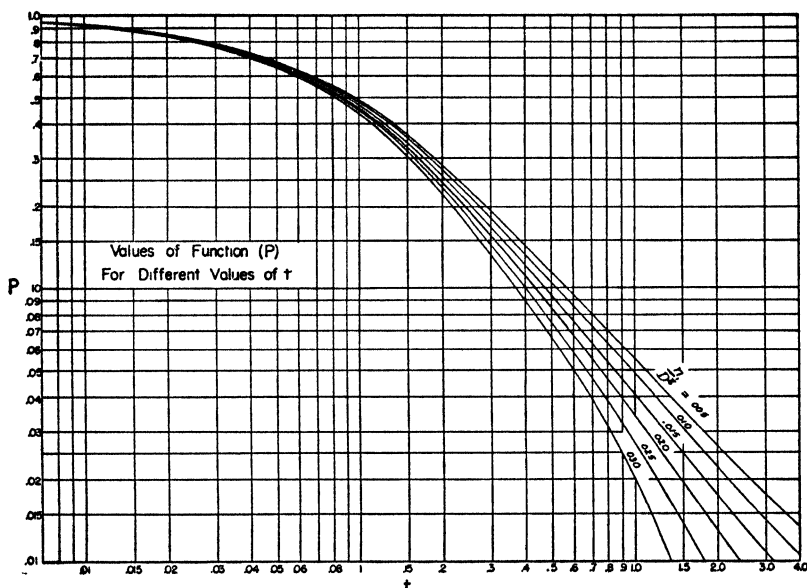


FIG. 4. Field data on relation between suspended material, bed composition, and hydraulic characteristics of the sediment and the river or canal.

use of Eq. 15 was 1.89 lbs. per second, and this result was obtained with the use of the diagram of Fig. 4 and a few simple calculations.

### C. RELATION BETWEEN SUSPENDED MATERIAL AND BOTTOM COMPOSITION

It should be recalled that in the above analyses suspended material transportation under equilibrium conditions was discussed. Of course, in actual practice this simplified condition may not always be approached, especially for the fine sizes of silt. The non-equilibrium condition is very complex and was briefly analyzed in a previous paper by the author (5). For equilibrium conditions the concentration of any sediment increases toward the bottom, and therefore it is necessarily logical to suppose that the volume concentration on the bottom is greater than in any other part of

the stream. If this is not so, then our theory of suspended material distribution is invalid.

The diffusion process due to the turbulence near a boundary, such as the bottom, is probably not similar in all respects to that in the main body of the stream. There is, of course, a transfer by the turbulence of material from the bottom into the main stream, otherwise all the suspended material would eventually settle out. The rate at which material settles out depends on the concentration of the suspended material near the bottom and its velocity of fall. Under equilibrium conditions this rate of settling must be equaled by the rate of transfer of material upwards from the bottom by the eddies. Thus for any size of material there must exist a relation between the concentration of the material in suspension near the bottom, its velocity of fall, the amount of this material in the bottom, and some parameter or parameters characterizing the turbulence.

It is assumed that the intensity of the turbulence is the important item which determines the ability of the turbulence to place material in suspension. Since experiments indicate that the velocity fluctuations tend to be statistically distributed according to the normal error law, it is possible to characterize the intensity of the velocity fluctuations by the root-mean square value of the fluctuations; that is, by the term  $\sqrt{\bar{v}^2}$ , where  $v$  is the transverse (upward) component of the velocity. The shear in turbulent flow is

$$\tau = \overline{\rho uv} + \mu \frac{dv}{dy}. \text{ However, when sediment is being placed in sus-}$$

pension the liquid is quite turbulent and therefore  $\mu \frac{dv}{dy}$  is small compared to the first term. If it can be assumed that the correlation in the bottom region between  $u$  and  $v$  is in general the same for various flow conditions, and that  $\bar{v}^2$  and  $\bar{u}^2$  maintain a definite relationship, then  $\tau_0 \propto \rho \bar{v}^2$ . Thus the bottom intensity of the turbulence  $\bar{v}^2$  is assumed to be proportional to  $\sqrt{\tau_0/\rho}$ , the so-called "friction velocity." The various assumptions in the analysis, of course, need experimental confirmation. Such data is being obtained, but no definite general results are as yet on hand.

If the value of  $\sqrt{\tau_0/\rho}$  characterizes the ability of the turbulence to pick sediment off the bottom, then a functional relationship can be set up between the suspended sediment concentration at the bottom,  $N_0$ , the bottom concentration of any size of material,  $N_b$ , the velocity of fall of the material,  $c$ , and  $\sqrt{\tau_0/\rho}$ . Such a rela-



tionship was first developed by Lane and the author (6) and it is as follows:

$$N_o/N_b = \phi(c/\sqrt{\tau_o/\rho}) \quad (16)$$

Note that  $c/\sqrt{\tau_o/\rho}$  is the parameter which we have been designating by  $t$  in previous analyses. If a functional relationship such as Eq. 16 exists, then a plotting of  $N_o/N_b$  against  $t$  should give a smooth curve. No laboratory data exists which permits such a plotting; however, there are some meager field data. Using the best field data available, a plotting was made as shown in Fig. 5, and though there is considerable scatter of the points, there does seem to be a strong tendency toward a grouping about a definite curve. Considering the inadequateness of the data and the widely varying conditions under which it was obtained by various individuals, the plotting in Fig. 5 seems quite significant. ( $N_o$  in Fig. 5 is in p.p.m. and  $N_b$  in per cent.)

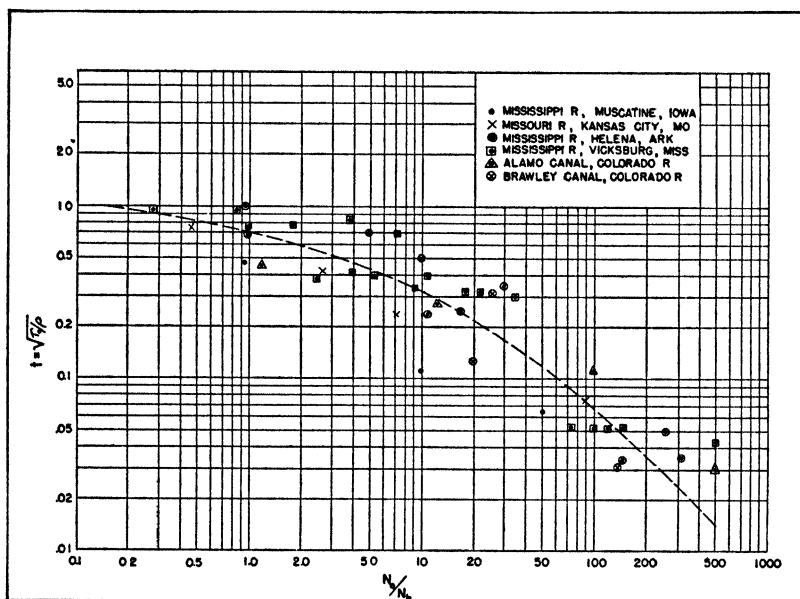


Fig. 5. Curves for obtaining values of function in the expression for total rate of suspended material transport.

The practical usefulness of a functional relationship such as shown in Fig. 5 is quite important. Such a relationship, after it has been definitely confirmed in the laboratory, permits the pre-

diction of whether under certain flow conditions the suspended load will be increased by scour or whether it will partly settle. It is noted that no material appears to be in suspension if the parameter  $t$  is greater than about unity. In other words, when the fall velocity of sediment in still water approaches the "friction velocity," transportation of such material in suspension becomes negligible.

## D. CONCLUSIONS

A study of the diffusion characteristics of the turbulence in a channel indicates that the so-called diffusion coefficient is almost zero at the bottom and the surface, and reaches a maximum at about the mid-depth. Theory indicates that this is true for infinitely wide channels; however, experiments in channels of finite width seem to verify this fact.

Experimental data check the theory of suspended sediment distribution based on the diffusion coefficient concept. Though this coefficient varies throughout any vertical section, the use of an average constant value for the coefficient gives a much simpler equation for the distribution of the suspended material. Field data for various large and small rivers indicate that this simpler relation is adequate for practical engineering use.

For equilibrium conditions there must exist some functional relationship between the bottom concentration of suspended material, velocity of fall of the material, bed composition, and the turbulence. Such a relationship is developed, and various field data seem to verify the existence of such a function.

The items relating to the turbulence in an open channel which merit further laboratory investigation are: (1) Relation between shear, bottom roughness, and intensity of turbulence in the bottom region. (2) Relative values of longitudinal and transverse velocity fluctuations, that is, the relation between  $\overline{v^2}$  and  $\overline{u^2}$ . (3) The intensity of the turbulence and the diffusion characteristics of the turbulence and the diffusion characteristics of the turbulence near a free surface.

In regard to suspended material the most important item meriting intensive laboratory study is the mechanism of the placement of sediment in suspension in the bottom region. Further analyses relating to the prediction of the absolute amount of suspended material under various conditions depends on an increased knowl-

edge of the diffusion characteristics of the turbulence in the bed region. A laboratory verification of the relationship given as Eq. 16 is imperative.

Though the analyses and studies described pertain more directly to the transport of suspended material by water, the fundamental principles should apply equally well for transport of dust and other settleable particles by air, or for the transport of any other type of solids by any turbulent fluid.

*Acknowledgment:*

The assistance of Mr. J. M. Robertson and Mr. Chung-Ling Pien, research assistants, in obtaining and working up the various laboratory experimental data is gratefully acknowledged.

*REFERENCES*

1. GILBERT, G. K. Transportation of Debris by Running Water, U. S. Geological Survey Professional Paper 86, 1914.
2. BAGNOLD, R. A. Transport of Sand by Wind, Geographical Journal, vol. 89, no. 5, 1937.
3. TAYLOR, G. I. Statistical Theory of Turbulence. Proc. of Royal Society of London, vol. 151A, 1935, p. 424.
4. ROUSE, HUNTER. An Analysis of Sediment Transportation in the Light of Fluid Turbulence. U. S. Department of Agriculture, Soil Conservation Service, T. P. 25, 1939.
5. KALINSKE, A. A. Suspended Material Transportation Under Non-Equilibrium Conditions. Trans., Amer. Geophysical Union, 1940.
6. LANE, E. W., AND A. A. KALINSKE. Relation of Suspended to Bed Material in Rivers. Trans., Amer. Geophysical Union, 1939.

UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

# Mass Transfer and Friction in Turbulent Flow

*By*

THOMAS K. SHERWOOD, Sc.D.\*

CHEMICAL engineering is probably more dependent on the science of fluid mechanics than any branch of engineering other than naval and aeronautical. The various important chemical engineering operations of heat transfer, distillation, drying, humidification, and gas absorption involve the interphase transfer of heat or material, or both. Both the chemical engineer and the industries employing these various transfer operations will benefit directly by the elucidation of the fundamental mechanism of interphase transfer of heat and material. Chemical engineering also involves the transport and flow measurement of all types of fluids, and chemical engineers were among the first to popularize the use of the Reynolds number as a basis for the correlation of pipe friction data.

The important operations involving mass transfer are four: transfer between a solid and a gas, between a solid and a liquid, between a gas and a liquid, and between two immiscible liquids. As an example we may consider the steady-state evaporation of water from a wet surface into a turbulent air stream. Near the liquid surface the mechanism of the transfer process approaches that of molecular diffusion, about which much is known. In the main turbulent air stream the process approaches that of eddy diffusion, about which relatively little is known. The over-all process can be evaluated only on the basis of knowledge of the quantitative contribution of molecular and eddy diffusion at various points in the gas stream, particularly in the region near the phase boundary.

The laws of molecular diffusion in gases have been developed from the kinetic theory, and have been carefully checked experi-

\* Associate Professor of Chemical Engineering, Massachusetts Institute of Technology.

mentally. For steady-state diffusion of one gas (water vapor) through a layer of thickness  $y$  of a second stagnant gas (air) we have the relation

$$N_A = \frac{DP}{RTy} \frac{\Delta p_A}{p_{BM}} = \frac{DP}{yp_{BM}} \Delta C \quad (1)$$

where  $N_A$  is the rate of diffusion of water vapor, g. mols/(sec.) (sq.cm.);  $D$  is the molecular diffusivity of water vapor in air, cm<sup>2</sup>/sec;  $P$  is the total pressure;  $\Delta C$  is the difference in concentration of water vapor in air, g. mols/cu.cm., at the two boundaries of the stagnant gas layer; and  $p_{BM}$  is the logarithmic mean of the partial pressures of air at the two boundaries of the layer. The product  $DP$  is independent of total pressure.

In the main turbulent gas stream transfer takes place predominantly by turbulent mixing, or eddy diffusion, which is much more rapid than molecular diffusion. By analogy to (1) we may write:

$$N_A = E \frac{dC}{dy} \quad (2)$$

but this relation is of little help, in that it serves only to define an "eddy diffusivity." E. Dryden (6) points out that  $N_A$  should be proportional to  $dC/dy$  for cases of diffusion over distances large compared with  $l_2$ , defined by Taylor as

$$l_2 = \int_0^\infty R_y dy \quad (3)$$

where  $R_y$  is the correlation between the values of the speed  $U$  (normal to the direction of diffusion) at two points distant  $y$  apart. If the distances are small compared with  $l_2$  the material diffuses at a uniform speed, and (2) does not apply.

Towle (17) has reported data on the spread of carbon dioxide and of hydrogen from a point source in a turbulent air stream. Equation (2) was found to apply, and the values of  $E$  obtained were very nearly proportional to the Reynolds number for the flow in the duct used. Hydrogen and carbon dioxide, differing 22-fold in molecular weight, diffused at the same rate, suggesting that  $E$  may be independent of the weight of the molecules or very fine dusts transported by eddy diffusion.

Woertz (14) has recently completed a study of steady-state transfer of water vapor across the 5.3-cm. gap between the flat

parallel walls of a large vertical rectangular duct  $5.3 \times 61$  cm. in cross section. Water evaporated continuously from a water film on one wall, and was absorbed by a film of strong calcium chloride solution on the opposite wall. The test section was preceded by a similar wetted-wall section to ensure steady-state conditions and minimize entrance effects. Hypodermic needle Pitot tube and sampling traverses were made to obtain the velocity pattern and the concentration gradient, respectively. The latter was found to be linear over the main central section of the duct, and values of  $E$  were obtained from Eq. 2 by dividing the rate of water vapor transfer by the slope of this curve representing concentration traverse. Helium, air, and carbon dioxide were used as the main gas stream. Velocity traverses and hot-wire turbulence measurements are also reported.

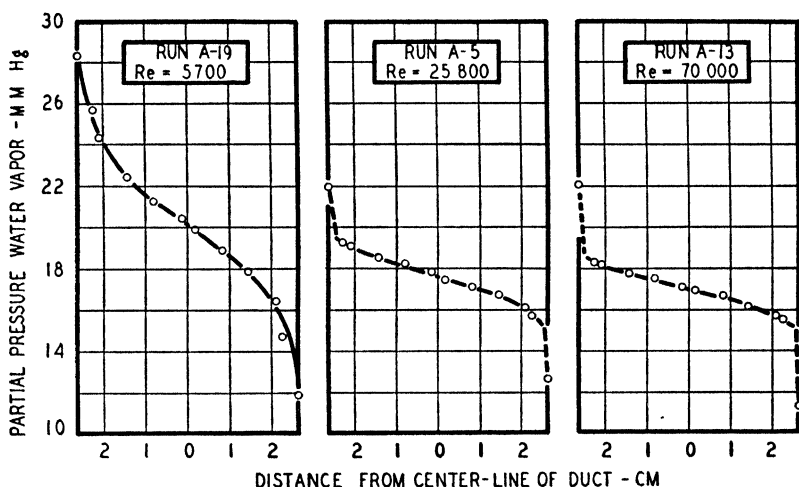


FIG. 1. Water-vapor concentration traverses for steady-state diffusion of water vapor across a turbulent gas stream between two flat walls 5.3 cm. apart.

Typical concentration traverses as obtained by Woertz are shown in Fig. 1, reproduced from the report of Woertz's results. The straight central sections, and the abrupt drops near the duct walls are particularly evident at the higher Reynolds numbers. Calculated values of  $E$  for air checked Towle's results very well, and the data for the three gases were correlated by plotting  $E_\rho$  vs  $Re$ , as shown in Figure 2.

Various writers have pointed out that  $E$  should be proportional to the product of a mixing length and the deviating velocity in

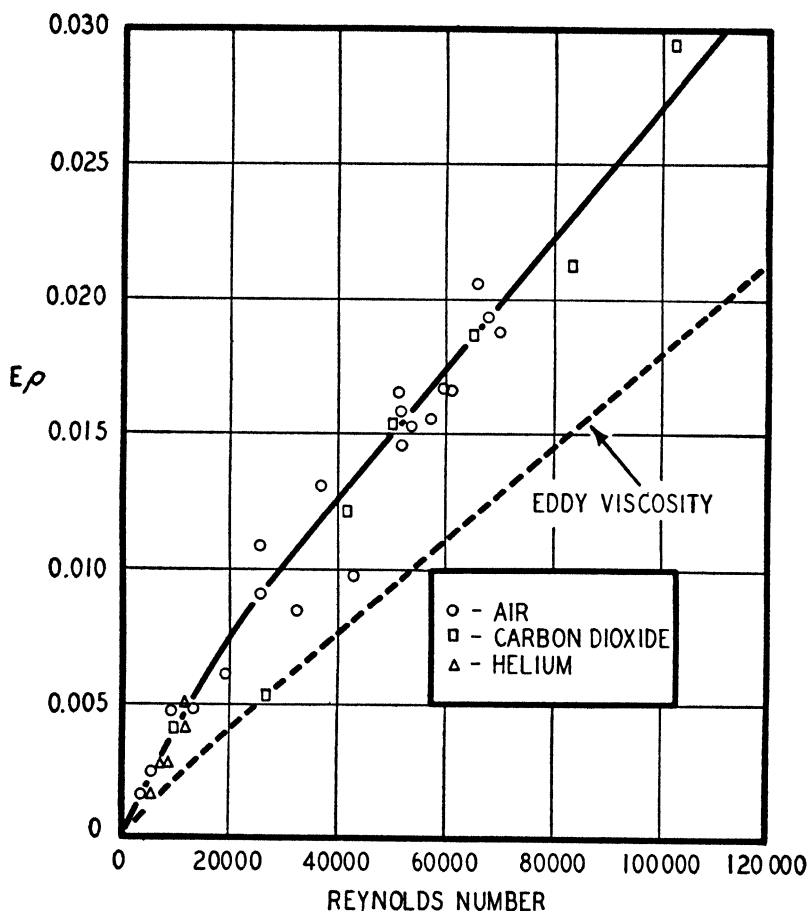


FIG. 2. Data of Woertz on eddy diffusion compared with eddy viscosity based on velocity traverses in the same duct.

the direction of diffusion. If the proportionality constant be  $\alpha$ , then

$$E = \alpha v' l \quad (4)$$

where  $v'$  is the root mean square deviating velocity in the  $y$  direction (direction of diffusion), and  $l$  is the Prandtl mixing length. Since the shear stress,  $\tau$ , is given by

$$\tau = \rho u'v' = \rho v'l \frac{dU}{dy} = \epsilon \frac{dU}{dy} \quad (5)$$

where  $\epsilon$  is the eddy viscosity, it follows that

$$E = \alpha v'l = \frac{\alpha \epsilon}{\rho} \quad (6)$$

Woertz's data, as shown in Fig. 2, indicate  $\alpha$  to be roughly constant at about 1.6, based on values of  $\epsilon$  obtained from velocity traverses and friction measurements in the same duct as employed in the diffusion tests.

### DATA ON MASS TRANSFER BETWEEN PHASES

A large amount of data are available on solution of solids, evaporation of liquids, gas absorption, etc., which may be compared with any theoretical equations proposed for mass transfer. As an example of such data we may describe briefly the results of Gilliland (7) on vaporization of several liquids into a turbulent air stream in a wetted-wall column. The liquid tested evaporated from a liquid film on the inner wall of a vertical tube 2.67 cm. i.d. and 117 cm. long. Nine different liquids were used, giving a four-fold variation of the molecular diffusivity,  $D$ . The Reynolds number for the air stream was varied from 1,840 to nearly 30,000 and the total pressure was varied from 0.14 to 3.07 atm. The results were interpreted in terms of Equation 1, the calculated values of  $B$  being taken as representing the thickness of a stagnant air layer which would have the same diffusional resistance as was actually encountered in the evaporation process. The data from a large number of tests were well correlated by the dimensionless equation

$$\frac{d}{B} = 0.023(Re)^{0.83} \left( \frac{\mu}{\rho D} \right)^{0.44} \quad (7)$$

The correlation obtained is indicated by Figure 3, where  $d$  is the pipe diameter.\* It follows from (1) and (7) that the rate of vapqr-

\* Gilliland later <sup>8</sup> obtained an equally good correlation of the same data by means of the equation

$$\frac{\Delta p_A}{N_A} = \left[ 903 + \frac{135}{D} \right] \frac{1}{Re^{0.8}} \quad (7a)$$

The two additive terms represent the relative resistances of core and film, respectively.



ization is proportional to  $D^{0.56}$ , which conclusion agrees qualitatively with the concept of transfer involving both molecular diffusion, for which  $N_A$  is proportional to the first power of  $D$ ,

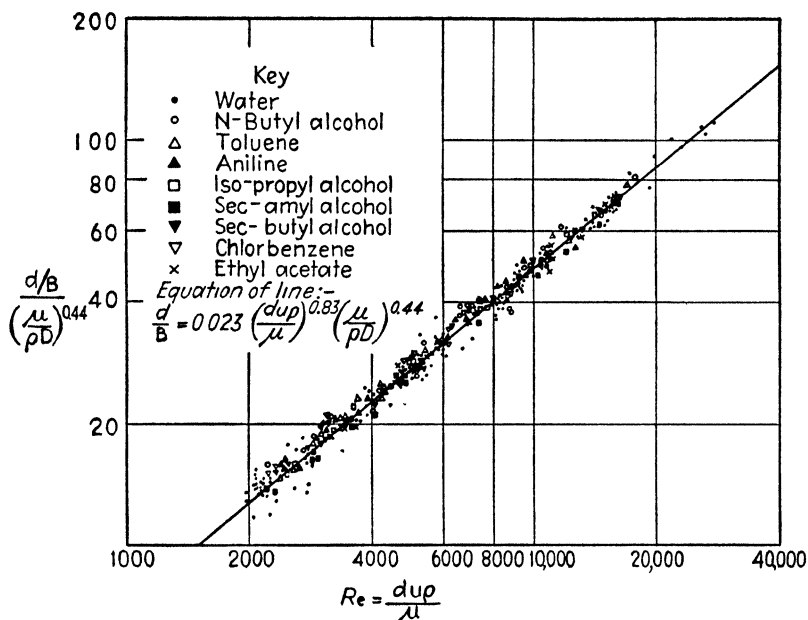


FIG. 3. Correlation obtained by Gilliland for vaporization of nine different liquids into a turbulent air stream in a wetted-wall column.

and eddy diffusion, for which  $N_A$  is proportional to the zero power of  $D$ . It also follows that  $N_A$  is inversely proportional to  $p_{BM}$ , as called for by Equation 1. This suggests that the rate of eddy diffusion is also inversely proportional to  $p_{BM}$ . In the tests described  $p_{BM}$  was nearly proportional to  $\rho$ , so this conclusion agrees with Equation 6 (the group  $\mu/\rho D$  is independent of pressure, and  $\epsilon$  is independent of pressure at a fixed value of  $Re$ ).

Gilliland's data were obtained in a standard tubular shape, with calming section, and cover a reasonably wide range of values of  $Re$  and  $\mu/\rho D$ . They may be taken as typical vaporization data, and will be used for comparison with the theoretical relations discussed below.

# THE REYNOLDS ANALOGY AND THE COLBURN ADAP- TATION OF THE PRANDTL-TAYLOR EQUATION

The well-known Reynolds analogy between heat transfer and friction is based on an assumption which may be expressed as follows: the head lost due to friction, divided by the momentum of the stream, is equal to the ratio of the actual heat transferred to the heat which would be transferred were the fluid to come to the temperature of the heat-transfer surface. This assumption, as applied to flow in round pipes, gives

$$h = \frac{1}{2} f c \rho U_o \quad (8)$$

where  $h$  is the heat transfer coefficient, g. cal/(sec.)(sq.cm.)(°C.),  $c$  is the specific heat of the fluid,  $\rho$  is its density,  $U_o$  its average velocity (cm.sec.), and  $f$  is the usual friction factor in the Fanning equation. The analogous assumption for mass transfer is: and head lost due to friction, divided by the momentum of the stream, is equal to the ratio of the actual material transferred to the material which would be transferred were the stream to come to equilibrium with the other phase. This is combined with a material balance and the Fanning equation to give (15)

$$\frac{N_A}{\Delta C} = k_c = k_g R T \frac{1}{2} f U_o \quad (9)$$

In order to compare with Equation 7, we may take  $\frac{p_{BM}}{P}$  as unity, and assume

$$f = 0.039(Re)^{-0.17}, \quad \text{whence (9) becomes} \\ \frac{d}{B} = 0.0195(Re)^{0.83} \left( \frac{\mu}{\rho D} \right) \quad (10)$$

This coincides with (7) exactly when  $\frac{\mu}{\rho D} = 1.34$ , but obviously makes no allowance for the rôle of molecular diffusion ( $D$  cancels out when (1) and (10) are combined). The analogy (9) fits the data on diffusion of certain vapors, just as the heat transfer analogy (8) fits data on air in pipes within about 10 per cent, but agrees very poorly with data on heating water.

A modification of the Reynolds analogy proposed by Prandtl

(13) and by Taylor (16) introduced the concept of a laminar film adjoining the phase boundary. It was assumed that heat passed through the laminar layer by molecular conduction, and the Reynolds analogy was applied only to the turbulent core of moving fluid. The resulting equation for heat transfer is

$$h = \frac{\frac{1}{2} f c \rho U_o}{1 - r + r \left( \frac{c \mu}{k} \right)} \quad (11)$$

where  $k$  is the thermal conductivity of the fluid, and  $r$  is the ratio of the fluid velocity at the boundary of the laminar film to the average velocity of the main stream. The thickness of the laminar film, and hence the value of  $r$ , is determined by the critical value of the "friction distance parameter"  $\gamma^+$ . If this critical value of  $\gamma^+$  be taken as 11.6, as given by Bakhmeteff (1), then  $r$  is given by the simple relation

$$r = 8.2 \sqrt{f} \quad (12)$$

Although (11) represents a definite improvement over the Reynolds analogy (8), it is well known to be a poor approximation to the data on heat transfer to liquids, especially oils.

The corresponding analogy for mass transfer has been derived by Colburn (4) as

$$k_c = \frac{\frac{1}{2} f U_o \frac{P}{p_{BM}}}{1 - r + r \left( \frac{\mu}{\rho D} \right)} \quad (13)$$

The relation of this equation to (9) is obviously similar to the relation of the Prandtl-Taylor equation (11) to the simple Reynolds analogy (8) for heat transfer. For low partial pressures of the diffusing gas  $p_{BM}$  and  $P$  are nearly equal, and the only change is the substitution of the group  $\frac{\mu}{\rho D}$  for  $\frac{c \mu}{k}$ .

A further modification of the Prandtl-Taylor equation has been suggested by Chilton and Colburn (3), who proposed the substitution of the empirical function  $\left( \frac{c \mu}{k} \right)^{2/3}$  for the denominator of (11), and the corresponding substitution of  $\left( \frac{\mu}{\rho D} \right)^{2/3}$  for the de-

nominator of (13). The two forms of the equation are the same, no matter what the value of  $r$ , for values of  $\frac{c\mu}{k}$  or  $\frac{\mu}{\rho D}$  of 1.0.

The modified forms may be written

$$j_H = \frac{h}{cU_o\rho} \left( \frac{c\mu}{R} \right)^{2/3} = \frac{1}{2}f \quad (14)$$

and

$$j_D = \frac{k_c p_{BM}}{U_o P} \left( \frac{\mu}{\rho D} \right)^{2/3} = \frac{1}{2}f \quad (15)$$

Colburn (5) has presented correlations of heat transfer data in the form of graphs of  $j_H$  vs.  $Re$  for several shapes, and Chilton and Colburn (3) have not only done the same for  $j_D$  in mass transfer, but have shown that  $j_H$  and  $j_D$  are approximately equal for the same shape at the same  $Re$ . This relation provides a valuable connection between mass transfer and heat transfer for a wide variety of conditions.

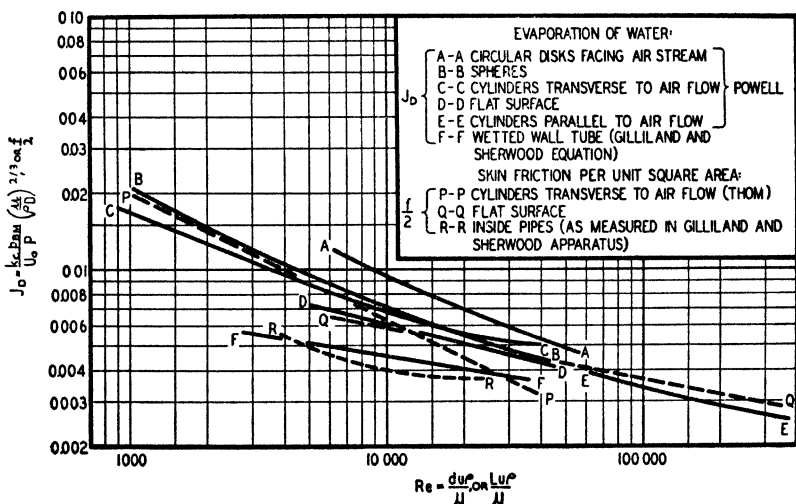


FIG. 4. Comparison of data on skin friction and on vaporization of water from wet surfaces of various shapes.

It is known that  $j_H$  approaches  $1/2 f$  for flow past streamline shapes, but that  $j_H$  may be only a small fraction of  $1/2 f$  for flow past vertical cylinders or bluff objects. It is evident that the relation between  $j_D$  and  $1/2 f$  must be similar, but it is of interest to compare the two functions quantitatively. As a basis for compar-

son, it is convenient to use the recent data of Powell (12) who reports data on vaporization of water from flat surfaces, cylinders, spheres, and disks placed in turbulent air streams. Other such data exist, but no other single investigator has reported data on a wide variety of shapes. The comparison is shown in Figure 4, which includes lines representing the vaporization data of Powell, together with the line FF representing Gilliland's data on vaporization of water in a wetted-wall tube. Powell's results are available only in the form of graphs, but the data appear to be well correlated by the co-ordinates of Figure 4, except in the case of spheres, where a slight spread of the data suggest an additional effect of diameter. The Reynolds number was based on the nominal diameter for the round shapes, and on the wetted length in the case of flat surfaces and cylinders placed parallel to the direction of air flow. The various curves fall fairly close together, the values of  $j_D$  for circular disks facing the wind being only about twice those for flow inside pipes. The data on cylinders placed parallel to the air flow represent an extension of the line for flat surfaces. The various shapes fall roughly in the order which might be anticipated, with the lowest curves representing the straight or streamline shapes.

In comparing vaporization and friction data it is necessary to distinguish between the coefficient  $C_D$  of total drag, and the skin friction coefficient  $f$ . The two are assumed to be equal for flow inside pipes and over flat plates, and  $\frac{f}{2}$  for these shapes is represented by the lines RR and QQ. The former is based on pressure-drop measurements in Gilliland's apparatus, and checks approximately with the standard curves for this shape. The line QQ for flat surfaces represents the relation

$$f = 0.074 Re^{-0.20} \quad (16)$$

as given by Goldstein (9). The line PP represents the skin friction for a transverse cylinder, as given by Thom and quoted by Goldstein (10). This last curve cannot be expected to be accurate since it is obtained by subtracting the normal pressure drag from the total drag, and  $f$  is small compared with  $C_D$ , particularly at high values of  $Re$ . No data were located on skin friction for spheres or disks facing the air stream.

The results shown in Figure 4 are summarized briefly in Table I for a value of  $Re$  of 15,000. The ratio of  $j_D$  to  $\frac{f}{2}$  is very nearly unity,

TABLE I  
*Comparison of Data on Vaporization and Friction at  $Re = 15000$*

| Shape   | $d$ or $L$<br>cm. | $j_D$    | $\frac{2j_D}{f}$ | $\frac{2j_D}{C_D}$ |
|---|-------------------|----------|------------------|--------------------|
| Circular disks facing air stream (area based on one face only). | 5 4-22 1          | 0 0078   | ....             | 0.014              |
| Spheres . . . . .   | 1 96-15.5         | 0.0061   | ....             | 0.12               |
| Transverse cylinders . . . . .                                  | 0 16-37 2         | 0 0060   | 1.17             | 0 032              |
| Flat surface parallel to stream . . . . .                       | 1.8-24 3          | 0 0055   | 1.02             | 1.02               |
| Cylinders parallel to stream . . . . .                          | 7 5-108 6         | (0 0055) | 1.02             | 1.02               |
| Inside pipes . . . . .  | 2 67              | 0 0043   | 1.14             | 1.14               |

not only for the streamline shapes but for transverse cylinders. On the other hand, the ratio of  $j_D$  to half the coefficient of total drag is seen to be much less than unity for transverse cylinders, spheres, and the disks.

## THE MURPHREE ANALYSIS

The Prandtl-Taylor concept of a laminar film at the phase boundary, with a wholly turbulent main stream, is now realized to be an over-simplification of the actual situation. Velocity traverses near the phase boundary point to a gradual change from laminar to wholly turbulent flow, and any successful theory of mass or heat transfer must be based on a reasonable guess or knowledge of the conditions in this boundary region.

Although Murphree's paper (11) presents a theoretical relation between friction and heat transfer in pipes, it is not difficult to revise the analysis to obtain the corresponding relation for mass transfer. Murphree assumes the eddy viscosity to be constant in the main turbulent core, but in the region near the wall is assumed to be proportional to the cube of the distance  $y$  from the wall. If this region or "film" in which  $\epsilon$  varies is assumed to have a thickness  $a-b$ , where  $a$  is the pipe radius and  $b$  is the distance from the center to the film boundary, then in the film

$$\epsilon = \epsilon' \left( \frac{y}{a-b} \right)^3 \quad (17)$$

where  $\epsilon'$  is the constant eddy viscosity in the main stream. The shear stress is given by

$$\tau = (\mu + \epsilon) \frac{dU}{dy} \quad (18)$$

and it follows that in the turbulent core the velocity distribution it is given by integration as

$$U_m - U = \frac{f\rho U_o^2}{4\epsilon'a} (a - y)^2 = c_1(a - y)^2 \quad (19)$$

By combining (17) and (18) and integrating, an equation is obtained for the velocity distribution in the film. The result is combined with (19) to obtain the velocity across the whole diameter, and it is found that the calculated velocity traverse agrees well with certain data of Stanton. Murphree then introduces Lees' equation for  $f$  in round tubes and Stanton's data for  $U_m/U_o$  in brass tubes, and so obtains a general relation for  $\epsilon'$  and film thickness in terms of  $Re$ . The results are given in the form of a table of values  $\phi$  and  $\frac{b}{a}$ , reproduced here as Table II. The ratio  $\frac{b}{a}$  is the ratio of

TABLE II  
*Variables Employed by Murphree*

| $Re$             | $f \times 10^3$ | $\frac{b}{a}$ | $\phi$ |
|------------------|-----------------|---------------|--------|
| 3000 . . . . .   | 11.16           | 0.7203        | 7.2    |
| 5000 . . . . .   | 9.64            | 0.8108        | 11.7   |
| 10000 . . . . .  | 7.94            | 0.8835        | 20     |
| 25000 . . . . .  | 6.24            | 0.9318        | 42     |
| 50000 . . . . .  | 5.28            | 0.9541        | 72     |
| 100000 . . . . . | 4.54            | 0.9700        | 121    |
| 300000 . . . . . | 3.66            | 0.9837        | 300    |
| 500000 . . . . . | 3.36            | 0.9875        | 470    |

the radius out to the film to the pipe radius, and  $\phi$  is the ratio  $\epsilon'/\mu$ . It is evident that the film thickness is determined by the intersection of two velocity distribution curves, that for the core being the parabolic velocity deficiency from (19) based on a constant  $\epsilon'$ , and that for the film being based on an assumed variation of  $\epsilon$  with distance from the wall.

In order to obtain the heat transfer coefficient, Murphree assumes the eddy conductivity to be  $c\epsilon$ , and integrates the conduction equation for heat transfer through the film. The coefficient  $h$  is defined as the heat flow through the film divided by the corre-

sponding temperature-drop. For mass transfer we may employ Equation (6) and carry out the same integration to obtain

$$\frac{D}{k_c d} = \frac{\left(1 - \frac{b}{a}\right) F(\psi)}{\left(1 + \frac{b}{a}\right) 3\psi^{1/3}} \quad (20)$$

where

$$F(\psi) = \frac{1}{2} \ln \frac{(1 + \psi^{1/3})^2}{1 - \psi^{1/3} + \psi^{2/3}} + \sqrt{3} \tan^{-1} \frac{2\psi^{1/3} - 1}{\sqrt{3}} + \frac{\pi\sqrt{3}}{6} \quad (21)$$

and  $\psi$  represents the dimensionless group  $\frac{\alpha\mu}{\rho D}$ .

By employing Table II in conjunction with (20) it is possible to obtain a relation between the three dimensionless groups  $\frac{D}{k_c d}$ ,  $\psi$ , and  $Re$ . Murphree shows this derivation for heat transfer to check remarkably well with various data on heat transfer to air and water, and to check reasonably well with data on heating oils. Woertz and the author (14) showed the partial pressure gradient at  $Re = 5700$  given on Figure 1 to agree fairly well with the curve calculated by the Murphree procedure, and erred in stating that the calculated film thicknesses became negative at high values of  $Re$ . Further comparison of this theory with mass transfer data will be presented after discussing Kármán's recent contribution.

## THE KÁRMÁN THEORY

Von Kármán (18) has proposed a modification of the Prandtl-Taylor theory, introducing experimental data on velocity traverses as a basis for the calculation of transfer in the region of the phase boundary. As before, the Reynolds analogy is applied in the main turbulent stream. In some ways his procedure is similar to that of Murphree, in that the degree of turbulence is assumed to vary gradually through the film.

Kármán's derivation was for the case of heat transfer in pipes, but it will be of interest to follow the derivation for mass transfer, and to compare the result with vaporization data. Assuming the relation (6), we may write

$$\mathcal{N}_A = - \frac{\alpha}{\rho} \left( \frac{\rho D}{\alpha} + \epsilon \right) \frac{dC}{dy} \quad (22)$$



In the turbulent region, or core, both  $\frac{\rho D}{\alpha}$  and  $D$  may be taken as small compared to  $\epsilon$ . Assuming further that the ratio  $\tau/N_A$  is constant, (22) and (18) may be combined and integrated to give

$$\frac{U_2 - U_3}{\tau} = \frac{\alpha(C_1 - C_2)}{\rho N_A} \quad (23)$$

where the subscripts refer to any two positions in the wholly turbulent region.

The shear stress  $\tau$  and the rate of mass transfer  $N_A$  are taken as constant through a "film" of thickness  $B_f$ , and the velocity at the film boundary,  $U_B$ , is obtained from (18) as

$$U_B = \tau_o \int_o^{B_f} \frac{dy}{\mu + \epsilon} \quad (24)$$

Similarly, from (22),

$$C_s - C_B = \frac{N_A \rho}{\alpha} \int_o^{B_f} \frac{dy}{\frac{\rho D}{\alpha} + \epsilon} \quad (25)$$

where  $C_s$  is the concentration at the wall, and  $C_B$  is the concentration at the film boundary. The variation of  $\epsilon$  through the film is based on the velocity gradient of Nikuradse (2), assuming the relation between  $y^+$   $\left( = \frac{y U_o \rho}{\mu} \sqrt{\frac{f}{2}} \right)$  and  $u^+$   $\left( = \frac{U}{U_o} \sqrt{\frac{2}{f}} \right)$  to be general, and applicable to other fluids and pipes. Taking

$$u^+ = U \sqrt{\frac{\rho}{\tau_o}} = f(y^+) \quad (26)$$

then

$$\frac{dU}{dy} = \sqrt{\frac{\tau_o}{\rho}} f'(y^+) \frac{dy^+}{dy} = \frac{\tau_o}{\mu} f'(y^+) \quad (27)$$

whence

$$\mu + \epsilon = \frac{\tau_o}{\left( \frac{dU}{dy} \right)} = \frac{\mu}{f'(y^+)} \quad (28)$$

Kármán arbitrarily assumes the film boundary to correspond

to  $y^+ = 30$ , and approximates  $f(y^+)$  between 0 and 30 by two simple functions:

$$\begin{aligned} u^+ &= f(y^+) = y^+; \quad f'(y^+) = 1 & 0 < y^+ < 5 \\ u^+ &= f(y^+) = 5 + 5 \ln \frac{y^+}{5}; \quad f'(y^+) = \frac{5}{y^+} & 5 < y^+ < 30 \end{aligned} \quad (29)$$

The comparison of these relations with Nikuradse's data may be

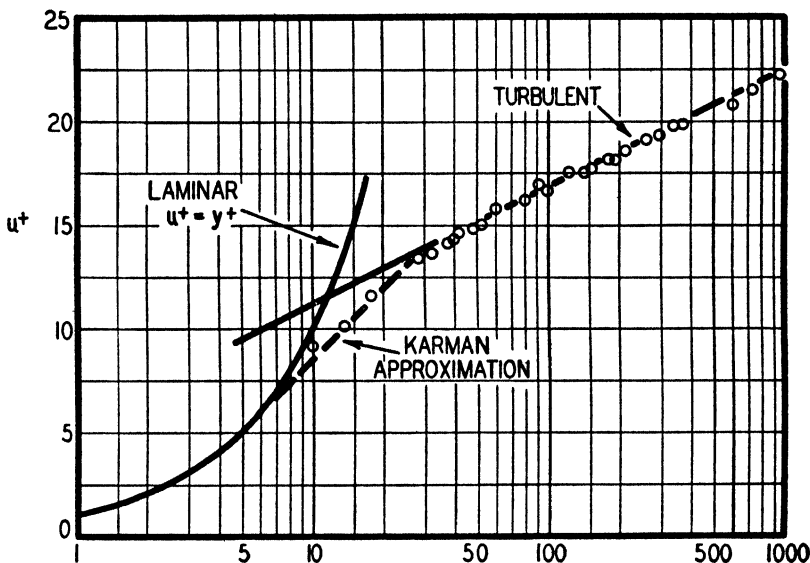


FIG. 5. Generalized velocity distribution, showing Nikuradse's data and approximation used by Kármán in region from  $y^+ = 5$  to  $y^+ = 30$ . (After Bakhmeteff.)

visualized by reference to Figure 5. The combination of (29) with (28) gives the required variation of  $\epsilon$ , so that (24) and (25) may be integrated to give

$$U_B = (5 + 5 \ln 6) \sqrt{\frac{\tau_o}{\rho}} \quad (30)$$

and

$$\frac{\alpha(C_S - C_B)}{N_A} = 5 \sqrt{\frac{\rho}{\tau_o}} [\psi + \ln(1 + 5\psi)] \quad (31)$$

where, as before,  $\psi = \frac{\alpha\mu}{\rho D}$ . Now if (23) be assumed to apply in

the turbulent region with the average concentration  $C_o$  replacing  $C_2$ , and the average velocity  $U_o$  replacing  $U_2$ , then

$$\frac{\alpha(C_B - C_o)}{N_A} = \sqrt{\frac{\rho}{\tau_o}} (U_o - U_B) \quad (32)$$

This is combined with (31) to give

$$\frac{\alpha(C_s - C_o)}{N_A} = \frac{\rho U_o}{\tau_o} + 5 \sqrt{\frac{\rho}{\tau_o}} [\psi + \ln(1 + 5\psi) - 1 - \ln 6] \quad (33)$$

whence

$$k_c = \frac{\frac{f}{2} \alpha U_o}{1 + 5 \sqrt{\frac{f}{2}} \left[ \psi - 1 + \ln \left( \frac{1 + 5\psi}{6} \right) \right]} \quad (34)$$

This result for round pipes may be seen to be of the same general form as the Reynolds analogy (9) and the Colburn equation (13).

In order to compare this adaptation of the Kármán theory with experimental data on mass transfer in pipes, we may take the empirical relation (7) representing Gilliland's tests. From (1) it is evident that if  $P$  and  $p_{BM}$  are taken as equal, then

$$\frac{d}{B} = \frac{N_A d p_{BM}}{\Delta C D P} = \frac{k_c d}{D}$$

whence

$$\frac{\alpha f U_o}{2 k_c} = \frac{\alpha f U_o B}{2 D} = \frac{f B}{2 d} R_e \psi \quad (35)$$

Substituting  $B$  from (7) and assuming  $f = 0.039 R_e^{-0.17}$ ,

$$\frac{f U_o}{2 k_c} = 0.85 \left( \frac{\psi}{\alpha} \right)^{0.56} \quad (36)$$

In order to compare with this empirical form representing the vaporization data, the Kármán form (34) is rewritten

$$\frac{f U_o}{2 k_c} = \frac{1}{\alpha} + \frac{5}{\alpha} \sqrt{\frac{f}{2}} \left[ \psi - 1 + \ln \left( \frac{1 + 5\psi}{6} \right) \right] \quad (37)$$

In making the comparison  $\alpha$  is taken as 1.6, the approximate value found by Woertz.

Figure 6 shows the results graphically with  $\frac{fU_o}{2k_c}$  plotted against  $\frac{\mu}{\rho D}$  for a Reynolds number of 10,000. The various theories indicate

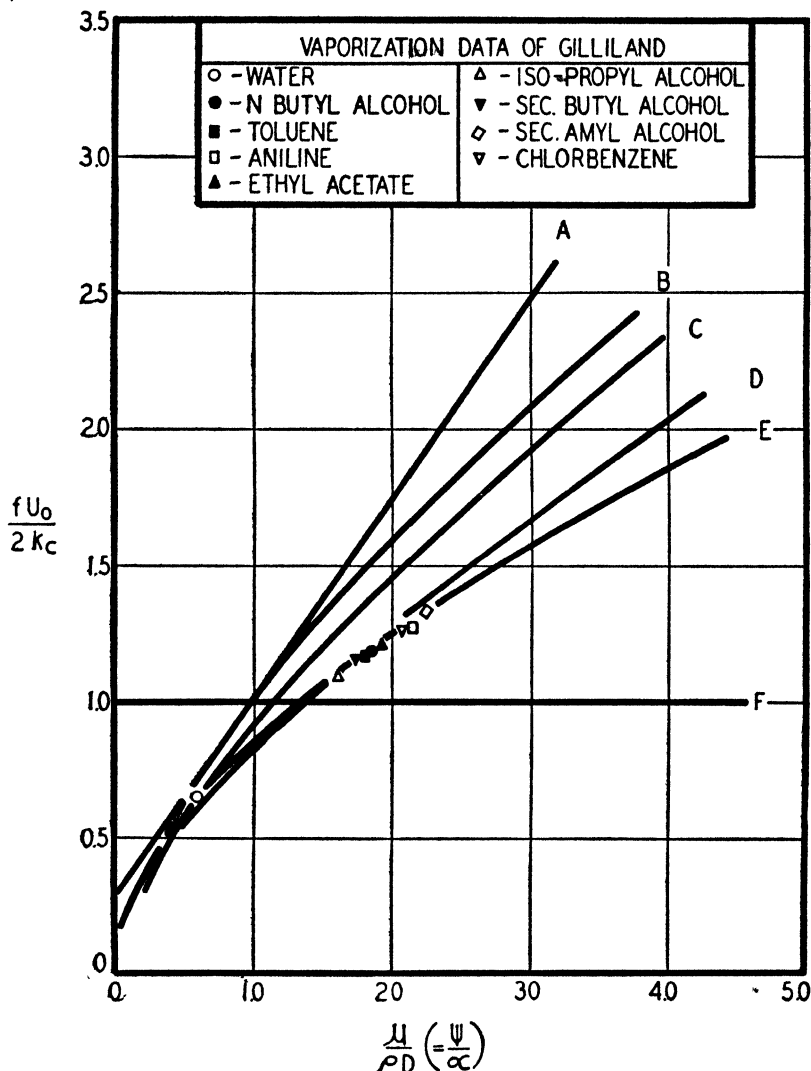


FIG. 6. Comparison of the various analogies with data on vaporization in a wetted-wall column at  $Re = 10000$ . Curves: A, Colburn, Eq. 13; B, Chilton-Colburn, Eq. 15; C, Murphree, Eq. 38; D, Kármán, Eq. 37; E, Gilliland empirical, Eq. 36; F, Reynolds analogy, Eq. 9.

relatively little variation of  $\frac{fU_o}{2k_c}$  with  $Re$ , so it is not surprising that Gilliland found no variation with  $Re$  (up to 30,000) of the exponent 0.44 on the group  $\frac{\mu}{\rho D}$ . The points represent the vaporization data for the nine liquids, and curve E represents Gilliland's empirical equation (36). The Kármán equation (37), represented by curve D, is seen to be in remarkable agreement with the experimental points. The maximum deviation is in the case of water, and is about 7 percent. The introduction of the factor  $\alpha$ , not used by Kármán, is obviously of considerable importance in bringing about this agreement. The linear equation given in a previous footnote as Eq. 7a would be represented by a straight line on Figure 6 passing through the experimental points. This form would agree more closely with the Kármán curve D than does the alternative empirical form represented by curve R.

Figure 6 also shows curves representing the various other theoretical equations. The Reynolds analogy leads to the horizontal line F, independent of both  $Re$  and  $\frac{\mu}{\rho D}$ . The Colburn equation (13) is represented by the straight line A, the position of which varies somewhat with  $Re$  due to changes in  $r$ . The Chilton-Colburn equation (15) is represented by line B, and is independent of  $Re$ . The curve representing the Murphree equation (20) is plotted by making the following rearrangement:

$$\frac{fU_o}{2k_c} = \frac{D}{k_c d} (Re)^{\frac{1}{2}} \psi = \frac{\left(1 - \frac{b}{a}\right) Re f \psi^{2/3} F(\psi)}{\left(1 + \frac{b}{a}\right) 6} \quad (38)$$

The result varies appreciably with  $Re$ , and curve C would be lowered about 15 per cent for  $Re = 50,000$ .

## EXTENSION OF THE KÁRMÁN ANALOGY TO STEADY-STATE TRANSFER

The remarkable agreement of the Kármán equation with the data, as evidenced by Figure 6, indicates that it is definitely superior to the other theoretical equations. In effect, Gilliland's extensive data on vaporization of various liquids have been predicted on the basis of Nikuradse's velocity traverses in water, the

known relation between  $f$  and  $Re$ , and a theory. It should be noted, however, that several questionable assumptions were introduced, and it may be asked if the errors due to wrong assumptions could have counterbalanced each other. It is first assumed that  $\mu$  and  $\frac{\rho D}{\alpha}$  are both small compared to  $\epsilon$ . This is certainly substantiated by the results of both Woertz and Towle, and cannot be seriously questioned. At  $Re = 20,000$  in air  $E$  and  $\epsilon$  are approximately 6.3 sq.cm./sec. and 0.043 g./(sec.)(cm.) respectively, so  $\mu$  is less than one half of one percent of  $\epsilon$ , and  $\frac{\rho D}{\alpha}$  is less than 4 percent of  $\epsilon$ . It is next assumed that  $\tau/N_A$  is constant in the turbulent core, whereas  $\tau$  is actually linear in  $y$  and the variation of  $N_A$  with  $y$  is different at different points along the pipe length. At the pipe inlet  $N_A$  is large near the wall and small in the core; near the outlet of a long pipe  $N_A$  is small at all values of  $y$ . In the analysis of the film  $\tau$  is assumed constant up to  $y^+ = 30$ , although at low flow rates this value of  $y^+$  may correspond to film thicknesses of half the radius, and the variation in  $\tau$  is actually 50 percent. Similarly,  $N_A$  is assumed constant in the film, although it must vary even more than  $\tau$ . In applying the Reynolds analogy (23) to the core, one limit was taken as the film boundary ( $y^+ = 30$ ) and the other as an indefinite point where  $U = U_o$  and  $C = C_o$ . This would appear to involve the assumption that these equalities occur at the same point. Finally, the general applicability of the relation between  $u^+$  and  $y^+$  is assumed, and an approximate representation of this relation is based on Nikuradse's data. This last must be accepted until better explorations have been made of the region near the phase boundary.

Many of the assumptions introduced by Kármán may be eliminated if the same general reasoning is applied to the derivation of a theoretical equation for steady-state transfer across a turbulent stream flowing between parallel plates. This was the case studied experimentally by Woertz, and his data may be used to check the relation obtained. For the core assume:

$$\mu + \epsilon = \frac{\rho D}{\alpha} + \epsilon = K = \text{constant} \quad (39)$$

and write

$$\tau = \tau_o \left( 1 - \frac{y}{a} \right) = K \frac{dU}{dy} \quad (40)$$

where  $a$  represents half the distance between the plates. Integration gives

$$U_m - U = \frac{\tau_o}{2Ka}(a - y)^2 \quad (41)$$

Similarly, from (22)

$$\frac{\alpha K}{\rho}(C_m - C) = -N_A(a - y) \quad (42)$$

The film is taken as the region from  $y^+ = 0$  to  $y^+ = 30$ ,  $u^+ = 0$  to  $u^+ = 13.9$ , whence

$$\frac{B_f}{a} = \frac{169}{Re\sqrt{f}} \quad (43)$$

where  $Re$  is defined as  $\frac{4aU_o\rho}{\mu}$ . From (42) and (43)

$$(C_m - C_B) = \frac{N_A\rho a}{\alpha K} \left( \frac{169}{Re\sqrt{f}} - 1 \right) \quad (44)$$

Combining (40) and (27) gives

$$\left( \frac{dU}{dy} \right) = \frac{K}{\left( 1 - \frac{y}{a} \right)} = \frac{\mu}{f'(y^+)} \quad (45)$$

so that the integration of (22) through the film takes the form

$$\begin{aligned} C_S - C_B &= \frac{N_A\rho}{\alpha} \int_0^{B_f} \frac{dy}{\frac{\rho D}{\alpha} + K - \mu} \\ &= \frac{N_A}{\alpha} \sqrt{\frac{\rho}{\tau_o}} \int_0^{30} \frac{dy^+}{\frac{\rho D}{\alpha\mu} + \frac{1}{f'(y^+)} - \frac{\mu}{\rho a} \sqrt{\frac{\rho}{\tau_o}} \frac{y^+}{f'(y^+)} - 1} \quad (46) \end{aligned}$$

This is integrated, as before, with the substitutions (29) to give

$$\begin{aligned} C_S - C_B &= \frac{N_A\rho a}{\alpha\mu} \ln \frac{1}{1 - \frac{5\psi\mu}{\rho} \sqrt{\frac{\rho}{\tau_o}}} \\ &+ \frac{N_A}{\alpha q} \sqrt{\frac{\rho}{\tau_o}} \ln \frac{(60p + 0.2 - q)(10p + 0.2 + q)}{(60p + 0.2 + q)(10p + 0.2 - q)} \quad (47) \\ \text{where } p &= -\frac{\mu}{5\rho a} \sqrt{\frac{\rho}{\tau_o}} = -\frac{1.13}{Re\sqrt{f}} \end{aligned}$$

$$\text{and } q = \sqrt{\frac{1}{25} + \frac{4.51}{Re\sqrt{f}} \left( \frac{1-\psi}{\psi} \right)}$$

The total concentration drop is  $2(C_s - C_m)$  and will be represented by  $\Delta C$ . It is obtained by combining (44) and (47) and rearranging, to give

$$\begin{aligned} \frac{U_o \Delta C}{N_A} &= \frac{Re}{2\alpha} \ln \frac{1}{1 - \frac{28.2}{Re\sqrt{f}}} - \frac{\mu}{2\beta a} \left( \frac{169}{Re\sqrt{f}} - 1 \right) \\ &+ \frac{2.82}{\alpha q \sqrt{f}} \ln \frac{(60p + 0.2 - q)(10p + 0.2 + q)}{(60p + 0.2 + q)(10p + 0.2 - q)} \end{aligned} \quad (48)$$

In this equation  $K$  has been replaced by  $\beta Re$ , since Woertz has shown both  $E$  and  $\epsilon$  to be nearly proportional to  $Re$ .

It is possible to compare this relation with Woertz's data, since he measured not only  $N_A$  and the concentration gradient, but noted the temperatures and concentrations of the liquid films on the walls of the duct. Thus  $\frac{U_o \Delta C}{N_A}$  calculated from his results does not employ his concentration traverse data, and his values of  $\beta$  and  $\alpha$  are not based on the values of  $C$  at the walls. For water in air  $\frac{\mu}{\rho D}$  is 0.60, and taking  $\alpha = 1.6$ ,  $\psi$  is 0.96. From Figure 2 the value of  $\beta$  is approximately  $0.018 \times 10^{-5}$ . As an approximation,  $\psi$  may be taken as 1.0 for both  $\text{CO}_2$  and air, and (48) reduces to

$$\begin{aligned} \frac{U_o \Delta C}{N_A} &= \frac{Re}{3.2} \ln \frac{1}{1 - \frac{28.2}{Re\sqrt{f}}} + \frac{8.8}{\sqrt{f}} \ln \frac{6 - \frac{169}{Re\sqrt{f}}}{1 - \frac{169}{Re\sqrt{f}}} \\ &+ 316 \left( 1 - \frac{169}{Re\sqrt{f}} \right) \end{aligned} \quad (49)$$

For helium  $\psi$  and  $\mu$  are approximately 3.0 and 0.00020, respectively, and an equation similar in form to (49) is obtained.

The comparison of these relations with Woertz's data, based on values of  $f$  as measured by Woertz, is shown on Figure 7. The two solid lines are the calculated curves for carbon dioxide and air, and for helium, based on  $\alpha = 1.6$ . The dotted curve is the calcu-



lated curve for air and carbon dioxide, with an approximate allowance for an increase of  $\alpha$  at low values of  $Re$ , as indicated by Figure 2. The general agreement is good with the exception of a group of points for air at high Reynolds numbers.

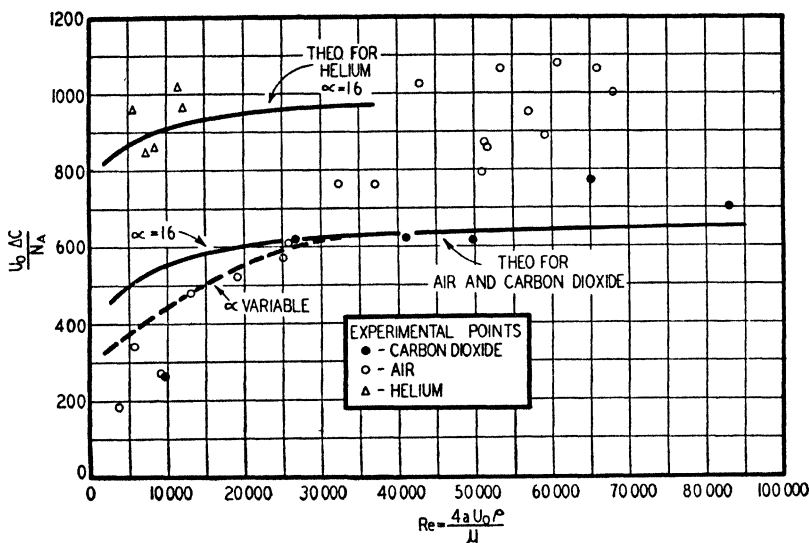


FIG. 7. Comparison of data on steady-state vaporization across a rectangular channel with revised Kármán analogy (Eq. 48).

In making this comparison it is important to visualize the nature of the experiment. Air passed vertically up through a rectangular passage  $5.3 \times 61$  cm. in section. A film of water covered one wall and a film of strong calcium chloride solution covered the opposite wall. The gas was recirculated, with a wetted-wall calming section below the test section, so that there was very little net transfer of water to the gas stream in passing through the test section.  $\Delta C$  was based on values of  $C$  corresponding to saturation at the temperature and vapor pressure of the main body of each liquid layer. Since there is doubtless some temperature and concentration difference between the surface and the main body of the liquid an error is introduced in  $\Delta C$ , varying both with liquid temperature and with  $N_A$ , or  $Re$ . These errors were not involved in Woertz's determinations of  $E$ , which were based wholly on  $N_A$  and the concentration traverse. The same type of error was present in Gilliland's work, but was presumably not nearly so large because the

Reynolds numbers were much lower and because pure liquids were vaporized, giving no possible concentration gradients as in the case of absorption of water by strong solutions of calcium chloride. Considering the difficulties of the experimental technique the agreement with the theory is quite good. Insofar as Kármán's principal assumptions were eliminated in deriving (48), the result may be considered as lending further support to the general concept on which Kármán's analogy is based.

### LITERATURE CITED

1. BAKHMETEFF, B. A. The Mechanics of Turbulent Flow, Princeton University Press, 1936.
2. BAKHMETEFF, B. A. Ibid., Fig. 51, p. 79.
3. CHILTON, T. H., AND A. P. COLBURN. Ind. Eng. Chem., **26**:1183, 1934.
4. COLBURN, A. P. Ind. Eng. Chem., **22**:967, 1930.
5. COLBURN, A. P. Trans. Am. Inst. Chem. Eng., **29**:174, 1933.
6. DRYDEN, H. I. Ind. Eng. Chem., **31**:416, 1939.
7. GILLILAND, E. R., AND T. K. SHERWOOD. Ind. Eng. Chem., **26**:516, 1934.
8. GILLILAND, E. R. Ind. Eng. Chem., **30**:506, 1938.
9. GOLDSTEIN, S. Modern Developments in Fluid Dynamics, p. 362, Clarendon Press, Oxford, 1938.
10. GOLDSTEIN, S. Ibid., p. 425.
11. MURPHREE, E. V. Ind. Eng. Chem., **24**:726, 1932.
12. POWELL, R. W. Inst. Chem. Eng. (London), Paper presented at meeting in London on March 8, 1940.
13. PRANDTL, L. Zeit. Physik, **11**:1072, 1910; **29**:487, 1928.
14. SHERWOOD, T. K., AND B. B. WOERTZ. Trans. Am. Inst. Chem. Eng., **35**:517, 1939.
15. SHERWOOD, T. K. Absorption and Extraction, McGraw-Hill Book Co., New York, 1937.
16. TAYLOR, G. I. A. C. A., Rep. and Mem., 272, 423, 1916.
17. TOWLE, W. L., AND T. K. SHERWOOD. Ind. Eng. Chem., **31**:457, 1939.
18. VON KÁRMÁN, TH. Trans. A. S. M. E., **61**:705, 1939.

### TABLE OF NOMENCLATURE

- $a$  = pipe radius or half clearance between parallel plates, cm.  
 $b$  = distance from center line to film boundary, cm.  
 $B$  = equivalent film thickness, cm.  
 $c$  = specific heat of fluid, g. cal/(g.)(°C.)  
 $c_1$  = a constant  
 $C$  = concentration, g. mols/cu.cm.

$C_B$  = concentration at film boundary, g. mols/cu.cm.

$C_D$  = drag coefficient = total drag per unit surface area  $/ \frac{1}{2} \rho U_o^2$

$C_m$  = concentration at centerline of channel, g. mols/cu.cm.

$C_s$  = concentration at phase boundary or pipe wall, g. mols/cu.cm.

$d$  = pipe diameter, cm.

$D$  = molecular diffusivity, sq.cm./sec.

$E$  = eddy diffusivity, sq.cm./sec.

$f$  = friction factor =  $2\tau_o / \rho U_o^2$

$h$  = heat transfer coefficient, g. cal/(sec.)(sq.cm.)(°C.)

$j_D$  = defined by first equality of Eq. 15.

$j_H$  = defined by first equality of Eq. 14.

$k$  = thermal conductivity, g. cal/(sec.)(sq.cm.)(°C./cm.)

$k_c$  = mass transfer coefficient, g. mols/(sec.)(sq.cm.)(g. mol/cu.cm.)

$k_G$  = mass transfer coefficient =  $\frac{k_c}{RT}$  = g. mols/(sec.)(sq.cm.)(atm.)

$K$  =  $\mu + \epsilon$ , or  $\frac{\rho D}{\alpha} + \epsilon$

$l$  = Prandtl mixing length, cm.

$l_2$  = Taylor mixing length, as defined by Eq. 3.

$L$  = length of wetted surface, cm.

$N_A$  = rate of mass transfer, g. mols/(sec.)(sq. cm.)

$p$  =  $-\frac{\mu}{5\rho a} \sqrt{\frac{\rho}{\tau_o}} = -\frac{1.13}{Re\sqrt{f}}$

$p_A$  = partial pressure of diffusing gas, atm.

$p_{BM}$  = logarithmic mean partial pressure of inert gas, atm.

$P$  = total pressure, atm.

$q$  =  $\sqrt{\frac{1}{25} + \frac{4.51}{Re\sqrt{f}} \left( \frac{1-\psi}{\psi} \right)}$

$r$  = ratio of fluid velocity at the boundary of the laminar film to average velocity  $U_o$ .

$R$  = gas constant, (cu. cm.)(atm.)/(g. mol)(°K.)

$R_y$  = correlation between two values of the speed  $U$ .

$Re$  = Reynolds number =  $\frac{dU_o\rho}{\mu}$  for pipes; =  $\frac{LU_o\rho}{\mu}$  for flat plates and

longitudinal cylinders; =  $\frac{4aU_o\rho}{\mu}$  for flow between parallel plates.

$T$  = absolute temperature, °K.

$u'$  = root mean square deviating velocity in direction of main flow, cm./sec.

$u^+$  =  $U \sqrt{\frac{\rho}{\tau_o}} = \frac{U}{U_o} \sqrt{\frac{2}{f}}$

$U$  = fluid velocity (time mean at a point), cm./sec.

$U_B$  = fluid velocity at film boundary, cm./sec.

$U_m$  = maximum or center-line velocity, cm./sec.

$U_o$  = average fluid velocity, volumetric flow rate divided by channel cross section, cm./sec.

$v'$  = root mean square deviating velocity in direction of diffusion, cm./sec.

$y$  = distance from phase boundary or pipewall, cm.

$$y^+ = \frac{yU_o\rho}{\mu} \sqrt{\frac{f}{2}}$$

$\alpha$  = proportionality constant in Eq. 4, dimensionless

$$\beta = \frac{K}{Re}$$

$\Delta C$  = driving force, or difference in concentration, g. mols/cu. cm.

$\epsilon$  = eddy viscosity, g./ $(\text{sec.})(\text{cm.})$

$\epsilon'$  = eddy viscosity in core, or main body of turbulent stream, g./ $(\text{sec.})(\text{cm.})$

$\mu$  = fluid viscosity, g./ $(\text{sec.})(\text{cm.})$

$\rho$  = fluid density, g./cu. cm.

$\tau$  = shear stress per unit area parallel to flow, dynes/sq. cm.

$\tau_o$  = shear stress at phase boundary or pipe wall, dynes/sq. cm.

$$\phi = \epsilon'/\mu$$

$$\psi = \frac{\alpha\mu}{\rho D}$$



UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

Contribution of Mathematical Statistics  
to Scientific Methodology

*By*

SAMUEL S. WILKS, PH.D.\*

I. INTRODUCTION

MODERN statistical methodology is the result of the application of mathematical concepts and methods to the description and the interpretation of sets of observations of a repetitive character which arise in scientific experiments, in mass production, distribution and consumption, in census work, and in many other fields. Statistical activity may be conveniently classified into two general categories. To one of the categories belong the routine collection, tabulation, description and presentation of masses of numerical data. This work has been greatly routinized within recent years by the use of machines; simple mathematical methods such as averaging, graphing, curve fitting, etc., are employed for purposes of condensing and presenting the data, and making simple comparisons. In the other category belongs a body of statistical methodology built up over the last forty years which now constitutes an important part of certain scientific and technological procedures. The methodology in this category has been developed for making predictions and inferences from observations together with appropriate measurements of confidence. Here we find the mathematical methods somewhat more advanced with the theory of probability playing a very fundamental rôle. In the present paper we shall confine our discussion to the principles of this class of methodology and their scientific significance.

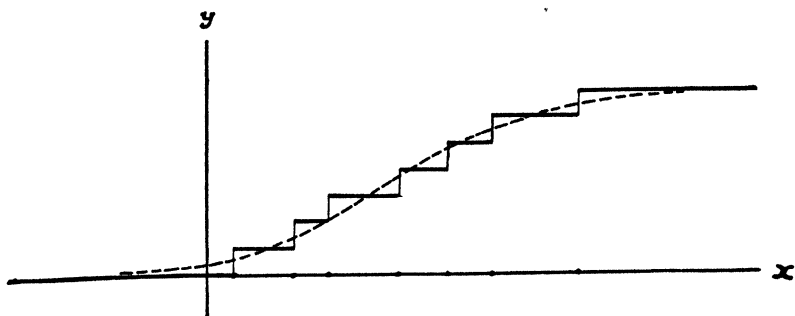
\* Associate Professor of Mathematics, Princeton University.

## II. NOTIONS AND PRINCIPLES OF PROBABILITY THEORY

Approaches ranging from formal mathematical ones to empirical ones have been made to the theory of probability, none of them entirely without objections. From the point of view of applied mathematics, probability theory has been devised for describing certain empirical phenomena. Probability theory is also recognized as an important field in pure mathematics. Perhaps the most realistic approach is to examine informally a simple empirical situation and see how we should proceed to build a "model" to describe the aspects of the situation in which we shall be interested. It will be found that a model can be constructed which can be used to advantage under certain conditions for describing and making calculations in problems dealing with repetitive phenomena in a manner analogous to the way in which geometry can be used in describing and making calculations in problems of surveying, construction, etc.

*An Example.* Suppose 30-watt electric light bulbs are made according to a certain set of specifications, and that we are interested in the length of life of the bulbs under continuous burning. If we let  $X$  denote the number of hours as determined under a given set of conditions that one of the bulbs will burn before expiring, then, of course,  $X$  will have different values for different bulbs. The sequence of bulbs coming from a production line is very similar in certain respects to a sequence of throws of a die. The variable  $X$  can be regarded as jumping from value to value as the bulbs emerge from the production line. In case of a die, we could let  $X$  denote the number of dots appearing on the uppermost face of a die. Then  $X$  would take on one of the values 1, 2, 3, 4, 5, 6, at each throwing of the die, and of course would jump from value to value as the die is successively thrown.  $X$  is referred to as a random variable and it is used to measure whatever property or quality possessed by each member of a sequence of repeated events we may be interested in. The collection of events generated by such a sequence may be referred to as a *population* of events. In the case of the bulbs the value of  $X$  cannot be determined except through the destruction of the bulbs. In view of this destruction, a study of the population of bulbs emerging from the production line would ordinarily be made on a basis of a "random sample"

taken from the line—random as far as  $X$  is concerned, although it may be taken by regularly choosing every 1000th bulb. Now if we test a sample of  $n$  cases, a certain proportion of them, say  $F_n(x)$  will have values of  $X$ , i.e., life lengths, less than or equal to  $x$ . This statement will hold for all values of  $x$ .  $F_n(x)$  is known as



the cumulative distribution function of  $X$  in the sample. The situation may be represented graphically as shown in Figure 1, where the  $n$  dots along the  $x$  axis represent the distribution of the  $n$  values of  $X$  in the sample, and the heavy step-like series of horizontal lines represent the graph of  $y = F_n(x)$ . If there is only one observation at each of the points shown on the  $x$  axis, then the jumps in  $F_n(x)$  will all have the same value, namely  $1/n$ . It is a matter of experience that if  $n$  were increased the graph of  $y = F_n(x)$  would under certain conditions appear to be approaching a smooth curve something like the dotted curve in Figure 1. Now what are these conditions?

*The idea of "randomness."* Before we could idealize and apply probability theory in connection with this example, we should have to obtain some assurance that no scheme of drawing a sub-sample from the actual sample of bulbs (the scheme of selecting a particular bulb depending only on information available before the value of  $X$  for the bulb is determined or used) would yield a cumulative distribution function, which, as  $n$  increases, appears to be "significantly" different from  $F_n(x)$ . Due to the difficulty of objectifying randomness and because of the practical necessity of having to deal with relatively small values of  $n$ , the problem of obtaining this assurance is, to a considerable extent, an empirical one relying on experience and sound judgment for its practical



solution. It has been found by Shewhart<sup>1</sup> and others who have applied statistical methods in quality control work that the very process of securing randomness or "statistical control" in a sampling sequence of products such as bulbs yields extremely valuable information in discovering sources of variation in quality, and consequently in determining what steps should be taken in the manufacturing process to achieve an increase in the level of quality in the units. For example, in the case of light bulbs, a study of the relationship between the burning life  $X$  and the information available on the bulbs before  $X$  is made available will frequently indicate which factors in the manufacturing process are associated with relatively short-lived bulbs and which ones are associated with relatively long-lived bulbs. Steps can then be taken to eliminate those factors associated with short-lived bulbs. A new sequence of bulbs may then be drawn and tested, repeating the process until it is no longer economically feasible to remove causes associated with short-lived bulbs.

The problem of establishing a state of statistical control in the sense used in quality control work is essentially a process of weeding out controllable causes of variation and securing greater homogeneity of product with respect to the variable under consideration. It should be pointed out, however, that there are situations in which well-defined finite populations of objects exist, as contrasted to those generated by processes or operations, which are *per se* important subjects of study by sampling methods. For example, the population of eligible voters of a given state may be sampled in order to get an estimate of its preference toward a given gubernatorial candidate; the population of readers of a certain magazine may be sampled in order to obtain an estimate of its opinion of certain editorial matter; the population of wheat farms in the United States may be studied through samples. In such situations as these, the main problem is to leave the populations as defined, however heterogeneous they may be, and devise methods of sampling which will yield sequences of drawings possessing the characteristic of randomness which we have already informally described.

*Probability Functions.* Once a state of statistical control or randomness has been established (largely by empirical methods) in the sampling sequence, one is in a position to idealize with prob-

<sup>1</sup> W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, D. van Nostrand and Co. (1931).

ability theory, for a state of statistical control or randomness simply means that the cumulative distribution function for any scheme of subsampling remains essentially unchanged regardless of how the subsample is selected from the sample, provided the individual items are not selected on a basis of their respective  $X$  values. We would then assume the existence of a *cumulative probability function*  $F(x)$  which may be described, as far as its connection with the actual observations are concerned, as though it were a function "apparently" being approached by the cumulative distribution function of the sample as a limit as  $n$  increases indefinitely.  $F(x)$  may be regarded for purposes of interpretation as the idealized proportion of objects in the population for which  $X = x$  and is called the probability that  $X \leq x$ . Note that we are making no attempt to formalize in a purely mathematical fashion the connection between the observational function  $F_n(x)$  and the assumed function  $F(x)$ . The formal mathematical discipline begins only when the existence of  $F(x)$  is assumed. It is clear that the assumed cumulative probability function  $F(x)$  would be required to satisfy the following conditions:  $F(-\infty) = 0$ ,  $F(+\infty) = 1$ ,  $F(x_1) \leq F(x_2)$ , when  $x_1$  and  $x_2$  are any two values of  $X$  such that  $x_1 < x_2$ . It will be observed that  $F(x_2) - F(x_1)$  is the probability that  $x_1 < X \leq x_2$ . In statistics, there are two especially important types of functions  $F(x)$  which may be described as follows:

(a) That in which  $F(x)$  is a continuous function which can be differentiated with respect to  $x$ , whose graph is an ogive-shaped curve of the type shown as the dotted curve in Figure 1. This case, known as the *continuous case*, arises in situations where the value of  $X$  is obtained from reading a continuous scale in terms of inches, pounds, etc.

(b) That in which  $F(x)$  is a step function whose graph is a series of horizontal lines of the type shown in the jumps occurring at isolated values. The jump occurring at a particular value of  $x$ , say  $x'$ , is the probability that  $x$  takes on the value  $x'$ . This is the *discrete case*, and pertains to situations in which the value of  $X$  is obtained by counting, or more generally where  $X$  can take on only certain isolated values in the continuum of real numbers.

In the continuous case and in view of the fact that  $F(x_2) - F(x_1)$  is the probability that  $x_1 < X \leq x_2$ , it is evident that  $\frac{dF(x)}{dx} dx = f(x)dx$ , say, is (except for terms of order  $(dx)^2$ ) the probability

that  $x < X < x + dx$ . The function  $y = \frac{dF}{dx}$  is known as the distribution function or elementary probability law of  $X$ . The graph of  $y = \frac{dF}{dx}$  is, in most statistical situations, an inverted bell-shaped type of curve, which lies above the  $x$  axis. The area between the  $x$  axis and the curve and to the left of the ordinate  $x = x'$  is  $F(x')$ .

In the discrete case, if we let  $p(x_i)$  be the jump of  $F(x)$  at  $x = x_i$ , then  $y = p(x_i)$  is the probability that  $X = x_i$ , and the graph of  $y = p(x)$  would consist of ordinates erected at the isolated "possible" values of  $X$  and 0 otherwise. The mean value of the random variable  $X$  is defined by taking the product of each value of  $X$  and the probability corresponding to this value of  $X$  and summing for all values of  $X$ . More precisely  $E(X) = \int_{-\infty}^{\infty} xf(x)dx$  or  $\sum x_i p(x_i)$  for the continuous and discrete cases respectively. More generally, the mean value of any function of  $X$ , say  $\varphi(X)$ , is given by  $E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x)f(x)dx$  or  $\sum \varphi(x_i)p(x_i)$ . In particular, the variance  $\sigma^2$  of  $X$  is defined by  $E[(X - E(X))^2]$ .  $\sigma^2$  is a very useful index and describes the amount of dispersion or spread in the distribution of  $X$ .

The extension of our discussion to situations in which two or more random variables are required to describe each element in the population is straightforward, but it will not be undertaken here because of lack of space.

The problem of measuring on the basis of the evidence furnished by a sample of size  $n$  the credibility of the hypothesis that  $F_n(x)$  is behaving as though it were converging to the assumed  $F(x)$  would have to be considered before we should know how much confidence could be placed in  $F(x)$  as the appropriate cumulative probability function for the population in the given situation. There are many sets of circumstances, such as are found in games of chance, genetics, etc., where one can calculate appropriate probability functions  $F(x)$  on an *a priori* basis by combinatorial methods with remarkable success, that is, one can by deductive methods arrive at functions  $F(x)$  which are approached very closely by the corresponding  $F_n(x)$  for practical values of  $n$ . However, in problems of a more highly statistical type, such as the light bulb problem, the determination of the appropriate  $F(x)$  is not subject to such clear-cut *a priori* procedure. The usual pro-

cedure here is to assume  $F(x)$  to be of a certain functional form which involves parameters or constants which have to be estimated for the particular problem at hand. The problem of deducing the functional form often involves a considerable amount of mathematical ingenuity, and keen insight into the subject matter of the problem.

The most commonly used cumulative distribution function is the normal or Gaussian form  $F(x) = \int_{-\infty}^x f(x) dx$  where  $f(x)$  is the well-known normal distribution function  $\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} (x - a)^2\right)$ .

The parameters are  $a$  and  $\sigma$ , the mean and the standard deviation respectively of the distribution. By determining suitable values of  $a$  and  $\sigma$ , it has been found that the normal distribution function is satisfactory for many practical problems in statistics.

### III. SAMPLING THEORY

At this point it should be made perfectly clear that probability theory as far as formal mathematics is concerned begins only when  $F(x)$  is given. Mathematical probability is concerned with the application of technical mathematical operations of various kinds to the idealized cumulative distribution function  $F(x)$ . The theory of sampling and significance tests in statistics are built up in this manner. In the theory of sampling, we are usually interested in some average or index or "statistic"  $S_n$  calculated from the values of  $X$  ( $x_1, x_2, \dots, x_n$ ) of a sample of  $n$  items or individuals. The value of  $S_n$  fluctuates from sample to sample and in a set of  $N$  repeated samples each of size  $n$ ,  $S_n$  has its own cumulative distribution function  $F_N(s_n)$  to which corresponds a cumulative probability function  $F(s_n)$ . The function  $F(s_n)$  is rigorously determined by technical mathematical methods from the underlying cumulative probability function  $F(x)$  which is assumed to exist for the population of objects from which the sample is assumed to be drawn. As a matter of fact,  $F(s_n)$  in the continuous case is obtained by integrating the product  $f(x_1)f(x_2) \dots f(x_n)$  over all sets of values of the  $x$ 's for which  $S_n \leq s_n$ , while in the discrete case  $F(s_n)$  is obtained by summing the product  $p(x_1)p(x_2) \dots p(x_n)$  over all sets of values of the  $x$ 's for which  $S_n \leq s_n$ . The field of theoretical sampling—the building of sampling distribution models of the type

$F(s_n)$ —has been developed at an extremely rapid rate by mathematical statisticians in England and in this country within the last quarter century.

The problem of relating probability theory to empirical situations is much more elusive and requires much more caution than that of building up the theory itself. The construction of the theory is a deductive process, while the task of applying the theory to statistical inference is a complicated mixture of trial and error, induction and deduction. The practical value of probability theory when its validity in a given statistical situation is established, lies in the fact that it furnishes an experimentally verifiable method of estimating the relative frequency with which it can be expected that the value of a random variable  $X$  will fall between any given pair of limits in repeated observations. The variable  $X$  may be a measurement on a single item or object; or more generally, it may be an average or some other index calculated from the individual  $X$ 's of a sample of  $n$  objects.

#### IV. STATISTICAL ESTIMATION AND SIGNIFICANCE TESTS

Sampling theory together with its application in statistical estimation and in the construction of significance tests constitutes a very significant contribution to scientific methodology. The simplest kind of situation in which sampling theory is used is that in which a population of objects exists and it is desired to estimate the value of some measurable characteristic of the population, for example the average of some variable or the proportion of the individuals possessing a certain trait. Polls such as those conducted by Gallup are examples of simple sampling methods for estimating the proportion of eligible United States voters in favor of a given issue or candidate. The type of sampling carried out in such work is known as representative sampling, and the method consists in dividing up the population into smaller and more homogeneous groups and sampling at random from these sub-groups, the sizes of the sub-samples being proportional to the numbers of individuals belonging to these sub-groups. Applications of representative sampling in this and more general forms are becoming more and more numerous in crop estimation, market research, national resource surveys, etc., where it is impracticable or even impossible

to make an exhaustive survey of the population under consideration.

A more general type of problem in estimation arises in situations where there are grounds for assuming the cumulative distribution function  $F(x)$  associated with the underlying population, to be of a given functional form depending on one or more unspecified parameters which have to be estimated from the sample. Thus if  $\theta$  is such a parameter,  $F(x)$  will be of a form  $F(x, \theta)$  let us say. A fundamental assumption here, of course, is that there exists a value of  $\theta$  which will make  $F(x, \theta)$  the actual cumulative probability function for the population under consideration. We may refer to such a value of  $\theta$  as the "true" value.

*The Method of Maximum Likelihood.* There are many ways of calculating estimates of  $\theta$  from a given sample. If we conceive of a repetition of samples of a given size  $n$ , then of course the estimate of  $\theta$  for the given method of estimation will fluctuate from sample to sample about the "true" value of  $\theta$  as a mean. This suggests that in constructing our mathematical model of sampling, we should try to find a method of estimation which, for a given sample size, would yield estimates subject to a minimum amount of fluctuation from sample to sample. Such a method which is applicable to large samples from populations having assumed underlying distribution functions of a fairly wide class was discovered by R. A. Fisher<sup>2</sup> in 1922. This method of estimation, now known as the *Method of Maximum Likelihood*, consists of determining the value of  $\theta$  which maximizes the probability of obtaining the set of values of  $X$  observed in the sample, that is, of maximizing  $f(x_1, \theta)f(x_2, \theta) \dots f(x_n, \theta)$  in the continuous case and  $p(x_1, \theta)p(x_2, \theta) \dots p(x_n, \theta)$  in the discrete case,  $f(x, \theta)$  being used to denote  $\frac{dF(x, \theta)}{dx}$

in the continuous case and  $p(x', \theta)$  to denote the jump (probability) at  $X = x'$  in the discrete case. The *Method of Maximum Likelihood* provides not only the "best" estimate of  $\theta$ , in the minimum variance sense, but also the limiting form of the distribution law of this estimate in large samples from which one can determine the probability of obtaining a sample in which the estimate of  $\theta$  deviates from the "true" value by more than any given amount. Numerous improvements and refinements have been made by

<sup>2</sup> R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," *Phil. Trans. Roy. Soc. London, Series A*, Vol. 222, 1921.

Fisher, Neyman, and others on the general problem of estimation since Fisher's original work was done. For example, the method of fiducial argument or interval estimation has been devised to provide in certain cases "observable" confidence limits of parameters, even for small samples, which will include between them with a given preassigned probability the "true" value of the parameter to be estimated.

The significance of estimation theory in the methodology of certain types of scientific work is fairly evident as far as fitting distributions is concerned. The value of estimation theory goes much farther than this, however.

*The "Student" Test of Significance in Experimental Work.* A great deal of the work going on in experimental science is concerned with the variation of experimental factors and making comparisons of observations obtained under the different sets of experimental conditions. To return to the example of the 30-watt electric light bulbs, suppose it is desired to test the effect of a new filament design on the life of the bulbs. A number of bulbs would be manufactured in which the new filament design would be used—this being the only essential change introduced. Suppose a sample of new bulbs is selected and their individual burning lives determined. Let  $\bar{x}$  be the average burning life of the sample bulbs. Suppose the average burning life in bulbs with "regular" filaments as determined from a large number of tests is  $a$ . The question which now arises is the following: Is  $\bar{x}$  "significantly" larger than  $a$ , or could the difference have "reasonably" arisen as a result of sampling fluctuations, under the assumption that the sample of new bulbs has been drawn from a normal population with the same average burning life as that for the bulbs with regular filaments? The answer to this question is furnished by sampling theory. Clearly, if the sample can be regarded as a not-too-improbable sample from a population which has the same average burning life as that of "regular" bulbs, then there are no grounds, on the basis of the sample of available information on the new filaments, for regarding the new filaments as providing longer burning life on the average. The usual statistical procedure for deciding whether or not  $\bar{x}$  in a sample of size  $n$  for example is significantly greater than  $a$ , assuming the sequence of new bulbs to be in a state of statistical control and that the burning lives of the bulbs are distributed according to a normal distribution function, is to set up

a "significance level," say .95, and calculate the two confidence limits

$$\bar{x} \pm \frac{t}{\sqrt{n}} s \quad \text{where } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ and } t \text{ is}$$

such that

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n-1} \sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right)} \int_{-t}^t \frac{dx}{(1+x^2/(n-1))^{n/2}} = .95.$$

Values of  $t$  for various values of  $n$  and various significance levels have been tabulated and widely spread through statistical literature. For example if  $n = 25$ ,  $t = 2.02$ , and as  $n$  increases indefinitely,  $t$  decreases to 1.96. This test was, in its essentials, devised in 1908 by a British industrial statistician who wrote under the name of "Student."<sup>3</sup> It can be shown by sampling theory that under "ideal" conditions of repeated sampling about 95% of the samples will be such that the mean burning time of the population of new bulbs will be included between their respective pairs of confidence limits. If the smaller of the confidence limits exceeds the average burning life  $a$ , of the regular bulbs, we conclude that  $\bar{x}$  is significantly larger than  $a$  at the .95 probability level, and that the filament is conducive to longer burning life on the average.

It is true, of course, that in the event that "practically all" of the sample values of  $X$  are larger than  $a$ , we should be inclined to conclude on the basis of common sense that we were getting longer burning life by using the new filament. Although many experiments yield clear-cut results of this type in which common sense can be trusted to detect a genuine difference between experimental results, it is true that many classes of experiments do not fall into this category. Examples are those experiments in agriculture, industry, market research, etc., where attempts are being made to discover ways and means of improving yield, quality, and performance, and where the improvements occur in very small intremments. Significance tests *when properly applied* will prevent, to within the significance level adopted, the error of interpreting a difference as genuine, when in fact it is essentially a difference due to sampling fluctuation. Of course there is a possibility that a significance

<sup>3</sup> "Student," "The Probable Error of a Mean," *Biometrika*, Vol. 6 (1908).



test will fail to reveal a genuine difference when it exists, but this source of error can be controlled to a large extent by obtaining assurance that the functional form of the distribution function  $F(x)$  and the associated significance test are appropriate and that a state of statistical control or randomness exists so that they can be applied.

*The Pearson Chi-Square Test of "Goodness" of Fit between Observations and Hypothesis.* We have discussed only one of many significance tests which have been proposed and used in practice for various types of problems. There are problems in which it is desired to test whether or not two or more samples (perhaps representing as many variations of experimental factors) could reasonably have originated from populations having identical distributions without caring what the distributions are. Again, we may wish to test whether or not the sample cumulative distribution function  $F_n(x)$  is sufficiently close to a completely specified distribution  $F(x)$  deduced from *a priori* considerations to justify the statement that the difference between  $F_n(x)$  and  $F(x)$  can be "reasonably" accounted for as a sampling fluctuation. One of the most widely used criteria for making tests such as these is the Pearson Chi-Square Test. For example, we may wish to test after 600 throws of a die which actually yield the frequencies 126, 96, 90, 84, 96, 108 for 1, 2, 3, 4, 5, 6 respectively whether the die is behaving like a "true" die. That is, are the observed frequencies significantly different from the "expected" frequencies (each equal to 100) to reasonably conclude that the behavior of the die is consistent with that of a true die? The appropriate criterion for making this test is the Pearson<sup>4</sup> Chi-Square criterion defined by

$$\chi^2 = \sum_{i=1}^6 \frac{(f_i - 100)^2}{100} = 12.1$$

where  $f_1 = 126$ ,  $f_2 = 96$ ,  $f_3 = 90$ , etc. It will be noticed that  $\chi^2$  is an index for measuring how far the actual frequencies depart from the "expected" frequencies—the greater the collective discrepancy, the greater the value of  $\chi^2$  and the smaller the collective discrepancy, the smaller the value of  $\chi^2$ . Under "ideal" conditions of sampling, it can be shown that, in repeated samples of 600 throws, less than 5% of the samples are expected to yield values of  $\chi^2$

<sup>4</sup>Karl Pearson, "On the Criterion that a given Set of Deviations from the Probable in the Case of Correlated Variables is Such that Can Reasonably be Supposed to have Arisen from Random Sampling," *Phil. Mag.*, Vol. 50 (1900).

larger than that observed in the present example, i.e., 12.1. We should conclude on the basis of the 600 throws that the die is behaving in a biased manner (aces too frequent) although there is a small probability (less than .05) that our exceptionally large  $\chi^2$  is being produced by a "true" die.

Problems in which the  $\chi^2$  test in one form or another is applicable in testing the "goodness of fit" between observation and hypothesis arise in almost every branch of science.

## V. ANALYSIS OF VARIANCE AND ITS SIGNIFICANCE IN EXPERIMENTAL WORK

There is a growing tendency in various fields of experimental science where statistical methods are applicable to make use of the principles of statistics in the actual layout and design of investigations before any data is collected so that the results of the experiment can be tested and interpreted in a fairly simple manner, and in many cases so that the basis for inference will be broader. By using suitably arranged experiments and by making proper application of a technique known as the *analysis of variance*, it has been found that one can in a single experiment obtain information for inferring the effects of each of several experimental factors on the characteristic being studied in the experiment. To take a simple example described by Tippett,<sup>5</sup> suppose that in a cotton mill it is desired to study the extent to which weight per lea (yarn count) of a cotton yarn fluctuates from bobbin to bobbin and from doffing to doffing for a given spindle, holding other factors such as the raw material, operators, etc., as constant as is economically feasible. The variation due to these factors can be studied simultaneously by selecting a lea from each of the five bobbins in a given spindle and repeating this several times (thus obtaining yarn from several doffings) for the same five bobbins. If five doffings are used we then have 25 leas, five coming from each of five bobbins and five belonging to each of five doffings. The variable in which we are interested is weight per lea. If we let  $x_{ij}$  be the weight of the lea from the  $j$ th bobbin in the  $i$ th doffing, we can arrange the data from this investigation in the following table:

<sup>5</sup> L. H. C. Tippett, "Some Applications of Statistical Methods to the Study of Variation of Quality of Cotton Yarn," Supplement, *Jour. Roy. Stat. Soc.*, Vol. 2 (1934).

Bobbin

|      |             |             |             |             |             |             |
|------|-------------|-------------|-------------|-------------|-------------|-------------|
|      | 1           | 2           | 3           | 4           | 5           | mean        |
| 1    | $x_{11}$    | $x_{12}$    | $x_{13}$    | $x_{14}$    | $x_{15}$    | $\bar{x}_1$ |
| 2    | $x_{21}$    | $x_{22}$    | $x_{23}$    | $x_{24}$    | $x_{25}$    | $\bar{x}_2$ |
| 3    | $x_{31}$    | $x_{32}$    | $x_{33}$    | $x_{34}$    | $x_{35}$    | $\bar{x}_3$ |
| 4    | $x_{41}$    | $x_{42}$    | $x_{43}$    | $x_{44}$    | $x_{45}$    | $\bar{x}_4$ |
| 5    | $x_{51}$    | $x_{52}$    | $x_{53}$    | $x_{54}$    | $x_{55}$    | $\bar{x}_5$ |
| mean | $\bar{x}_1$ | $\bar{x}_2$ | $\bar{x}_3$ | $\bar{x}_4$ | $\bar{x}_5$ | $\bar{x}$   |

Doffing

where the  $\bar{x}_i$  at the side are doffing means and those at the bottom are bobbin means, while  $\bar{x}$  is the mean of all 25 cases. We can set up as indices of inter-bobbin and inter-doffing variation the sums of squares of residuals  $S_b = 5 \sum_1^5 (\bar{x}_i - \bar{x})^2$  and  $S_d = 5 \sum_1^5 (\bar{x}_i - \bar{x})^2$ .

The total variation in the data is  $S = \sum_{j=1}^5 \sum_{i=1}^5 (x_{ij} - \bar{x})^2$ , and the residual or "error" variation after inter-bobbin and inter-doffing variation have been eliminated is  $S - S_b - S_d = S_e$ . Genuine variation in yarn count due to differences in bobbins will be reflected by excessively large values of  $S_b$  as compared with  $S_e$ ; a similar statement holding for doffings. The question now arises: When can  $S_b$  or  $S_d$  be regarded as excessively large? This question can be answered by sampling theory under certain assumptions. The assumptions are that (a) each  $x_{ij}$  can be regarded as being the sum of a general mean  $m$ , a constant  $r_i$  due to the  $i$ th doffing and a constant  $c_j$  due to the  $j$ th bobbin, and an "error" residual  $\xi_{ij}$ ; (b) the  $\xi_{ij}$  are all distributed according to the same normal law having  $a = 0$ , but unknown variance  $\sigma^2$  and (c)  $\sum_i r_i = \sum_j c_j = 0$ .

To test whether or not there is excessive variation from bobbin to bobbin we make the assumption that the  $c_j$  are all 0, and compare  $S_b$  and  $S_e$ .  $1/4 S_b$  and  $1/16 S_e$  can both be regarded as independent estimates of the same unknown quantity, namely the variance  $\sigma^2$  of the population to which the  $\xi_{ij}$  are regarded as belonging. Under the assumptions that the  $c_j$  are all zero, the sampling distribution of the ratio  $F = \left( \frac{1}{4} S_b \right) / \left( \frac{1}{16} S_e \right)$  is known, from which, for a given probability level, it can be determined whether  $S_b$  is

too large as compared with  $S_e$  to lend credibility to our assumption that the  $c_i$  are all zero. If  $S_b$  is found to be significantly large in relation to  $S_e$ , we refute the assumption that the  $c_i$  are all zero, and conclude that there is a genuine inter-bobbin variation. A completely analogous procedure can be carried out for determining whether or not the inter-doffing variation can be considered genuine. It should be remarked that one can test the plausibility of the additive assumption (a) by duplicating the experiment. The significance of establishing sources of variation in this manner as a step in certain kinds of problems in science and technology is self-evident.

It is possible to design the experiment so as to get information regarding the existence of further sources of variation of weight per lea by using more elaborate layouts. The analysis of variance is a comparatively recent development due largely to R.A. Fisher<sup>6</sup> and his followers, and it has rich possibilities in scientific investigation in many fields when properly applied.

## VI. SUMMARY

There are two main categories of statistical activity: one deals with the routine collection and presentation of masses of statistical facts, and the other is concerned with a methodology founded on probability theory for making inferences and predictions from observations. The present paper deals with the latter category. Notions and principles of probability theory and their application to problems of sampling, statistical estimation, and significance tests are developed. The significance of sampling theory, statistical estimation, and significance tests in scientific and technological procedures is described and illustrated by examples. The value of laying out various types of experiments and investigations according to certain statistical principles is briefly discussed.

<sup>6</sup>R. A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh.



UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

Contribution of Statistics to the Science  
of Engineering

*By*

WALTER A. SHEWHART, PH.D.\*

INTRODUCTION

THE potential contributions of statistics to the science of engineering are an important national asset; an asset of interest to all of us because it makes possible the most efficient and effective use of natural resources and human effort to satisfy human wants; an asset, however, that has for years remained frozen and is only now beginning to be utilized; an asset that can be used to the full only when engineers and others learn to use it as they have learned, in the past, to use the product of the scientist.

Much has appeared in the literature to indicate some of the contributions of statistics to date in the field of engineering and manufacturing. My object is not so much to review what has been done as to survey the potential contributions of statistics to the science of engineering. In doing this, I shall follow the old advice that the easiest way to reach the top is to go to the bottom of things, and I shall go to the bottom of the difference between engineering with, and without, statistics.

Let us recall how the applied scientist has wrought so many wonders for you and me to enjoy. He has done much of this with a comparatively simple but extremely powerful tool, namely, scientific method based upon the concept of physical laws of nature that assume perfect or certain knowledge of a set of facts and then state exactly what will happen at any future time. This method consists of three essential steps: hypothesis, experiment, and test of hypothesis. The fundamental difference between engineering with and without statistics boils down to the difference between the use of a scientific method based upon the concept of laws of nature that do not allow for chance or uncertainty and a scientific method based upon the concept of laws of probability

\* Research Statistician, Bell Telephone Laboratories.

as an attribute of nature. When viewed in this way, the potential contributions of statistics become quite simple indeed. Statistically scientific method is in fact a fundamental discipline that includes all of the customary scientific method based upon the concept of exact laws as a limiting case in which repetitive operations of any given kind always give identically the same results.

All that I shall try to do here is to sketch in broad outline how the statistician has helped to refine scientific method by overhauling and reworking each of its three fundamental steps and then to indicate briefly how this new tool can be used by the engineer. In place of hypotheses based on exact laws, the new scientific method introduces statistical hypotheses. The introduction of statistical hypotheses makes it necessary to conduct the experiments in such a manner that the statistical hypotheses may be tested, and the new method provides means of testing the statistical hypotheses with which it starts.

Needless to say, scientists have long realized that their discovered "laws" do not always fit observed phenomena exactly. All that they have claimed is that the scientific method based upon the concept of exact laws has enabled them to make remarkable progress in understanding the world and to attain knowledge that could be used by engineers and applied scientists. Deviations from the assumed exact laws were simply dismissed as errors by the pure scientist and allowed for in factors of safety by the engineer. However, we are now coming to realize that many of the errors of the pure scientist and factors of safety of the engineer are more properly designated as factors of ignorance. We are beginning to see that we must refine our scientific methodology if we are to minimize these factors of ignorance in the scientific explanations of the world and if we are to make the most efficient and economic use of natural resources in the development of things to satisfy human wants. Statistically scientific method provides the scientist with an improved tool by which to extend his knowledge, and the engineer with a means by which to extend his useful service to mankind.

## BASIC ENGINEERING PROBLEM

The engineer's job is to devise and develop the operations that, if carried out, will produce things that people want.<sup>1</sup> To do this,

<sup>1</sup> This is also the job of the applied scientist in the development of ways and means of making everything that we use.

he must be able to make things that have quality characteristics lying within previously specified tolerance ranges. Hence, a basic engineering problem is to devise an operation of using raw and fabricated materials that, *if* carried out, will give a thing wanted. The specified tolerance ranges for the quality characteristics of the thing wanted define a target for the engineer. He devises an operation and predicts that, if carried out, it will hit the target, but, since he does not have certain or perfect knowledge of facts and physical laws, he cannot be certain that a given operation will hit its target; in fact the best that he can hope to do is to know the *probability* of hitting the target. Here then is one fundamental way in which probability enters into everything that an engineer does.

Furthermore, if the thing produced fails to meet tolerance requirements, the engineer is penalized in one way or another. For example, if the quality of any piecepart fails to meet its tolerance requirements, a loss is incurred through rejection or modification of the defective part; if the time-to-blow of a protective fuse fails to meet its tolerance range, loss of property and even loss of life may result; if the time-to-blow of a fuse in a shell fails to meet its tolerance range, the shell may burst prematurely and kill members of the gun crew and, in any case, the round of ammunition will fail to fulfill its function of destruction within the ranks of the enemy. This means that when the engineer undertakes to use probability theory it is essential that he thoroughly understand the conditions under which its use will lead to valid predictions.

In what follows we introduce the term *operation* to include *any* experimental procedure designed to produce a previously specified result. In this sense, a production process is an operation, and a method of measuring is also an operation. Furthermore, an engineering operation or a production process may almost always be broken down into component operations. It should also be noted that even if only one thing of a kind is to be made, the operation devised by the engineer for producing this one thing is presumably capable of being repeated again and again so that the one thing to be produced may be thought of as but one of a class of an indefinitely large number of things that might be produced by repeating the operation again and again at will under the same essential conditions. In this way we may reduce the basic engineering problem of devising an operation to hit previously specified tolerance ranges to one that can be treated statistically, in that



statistical theory treats of the properties of certain kinds of repetitive operations.

To illustrate, let us consider a simple example in which only one quality characteristic of a thing is specified, let us say the length of a piecepart or the time-to-blow of a fuse. Let us symbolize this quality characteristic by  $X$ . The operation for producing a thing of quality  $X$  within specified limits,<sup>2</sup> if repeated again and again, would give rise to an indefinitely long sequence of values of quality  $X$  if taken in the order in which the things were produced, and these may be represented symbolically as follows:

$$X_1, X_2, \dots, X_i, \dots, X_j, \dots, X_n, X_{n+1}, \dots, X_{n+k}, \dots \quad (1)$$

Hence, every operation may be characterized not only by a word description but also by the characteristics of the potentially infinite sequence (1) corresponding to an indefinitely large number of repetitions. In fact, we shall again and again make use of the characteristics of such a sequence in what follows. If an operation is developed to produce only one thing of a kind, then the engineer is interested only in the first term of this potentially infinite sequence (1) but if the engineer is interested in developing an operation or production process to turn out an indefinitely large number of things of the same kind, then he is interested in all terms in this sequence.

In what follows, we shall try to see how statistics can help the engineer to solve his basic engineering problem of developing an operation that, if carried out, will produce an object with qualities that lie within previously specified tolerance limits.

## BASIC CONTRIBUTION OF CLASSICAL STATISTICAL THEORY

*Basic Statistical Hypothesis.* As a background for viewing the contribution of statistics to the solution of the basic engineering problem, we may state the fundamental hypothesis of applied classical statistical theory in the following way:

*Hypothesis I. Some repetitive operations exist that obey laws of probability. These are called random. The probability that such a random operation will give a previously specified event, as for example, the occurrence of a value of  $X$  within a previously specified tolerance range, is a definite number associated with that event.*

<sup>2</sup> Or any operation of measurement of some objective quality characteristic.

If we know the law of probability or chance that controls a given operation, we may use the mathematical distribution theory of the statistician to describe how statistics of samples of size  $n$  given by successive repetitions of such an operation will be distributed. Likewise, if we know that an observed sample has been given by a random operation, the statistician has established valid rules of procedure for using the sample as a basis for estimating the parameters in the law of probability underlying the random operation that gave the sample. In other words, the applied mathematical statistician has provided us with rules for making valid predictions if we know the law of probability and with efficient rules of discovering the functional form of such a law, including the values of the parameters, if we simply know that it exists.

Of course, the work of the mathematician is purely formal; for example, what he calls the operation of drawing samples of size  $n$  at *random* consists essentially of acting upon some given mathematical law of chance or distribution function in accord with previously specified mathematical rules. Hence, if such a statistical hypothesis is to be of any value in engineering or applied science, it is necessary to know what is meant in an operationally verifiable manner by drawing at random. This necessitates our study of the second or experimental step in a statistically scientific method.

*Basic Statistical Experiment or Operation of Drawing at Random.* Unless an experimentalist knows what it means to draw a sample at random, he is not in a position to make use of statistical hypotheses because he cannot get the data with which to make valid tests of them and, having accepted a statistical hypothesis as valid, he does not know what kind of events he can predict with validity. Without such knowledge, he would be somewhat like a physicist who knew all about mathematical physics but did not know how to distinguish and measure the physical properties appearing in his equations. For example, an engineer may wish to select a random sample of 50 pieces of a new kind of product from the first 1,000 pieces produced. Very often he will ask under such conditions if he can take every twentieth piece as it is produced. Sometimes the engineer will propose other schemes but, in general, it is found that none of those proposed can be used with much assurance of giving a random sample. Hence the starting point in the use of statistical theory is a clear understanding of what is to be taken as the meaning of a random operation.

Let me begin with a description of two operations that I shall

choose to call random.<sup>3</sup> Let us assume that we have  $N$  physically similar chips on each of which is written a number and that these  $N$  chips are placed in a bowl. A blindfolded experimentalist thoroughly mixes the chips in the bowl, draws a number, and has his assistant record the number. The chip is then returned to the bowl and the blindfolded experimentalist, after thorough mixing of the chips, again draws one and has the number recorded. This random operation of drawing a number can theoretically be repeated again and again under the same essential conditions so as to give an infinite sequence like (1).

The other important operation for our present study is the following one of randomizing a finite set of  $N$  numbers. In this case the blindfolded experimentalist follows the same procedure as described above except that he does not return a chip to the bowl after it has been drawn. The  $N$  numbers drawn in this manner may be written down by the assistant in the order drawn. The operation of putting the  $N$  chips in a bowl, thoroughly mixing them, drawing them one at a time, and writing the numbers down in the order observed, may be repeated again and again at will so that the result of the operation of drawing one sequence is but one of the infinite number of sequences that might be obtained by repeating the operation again and again.<sup>4</sup> As an example, Fig. 1a records one such random drawing of the 144 values of thickness of inlay given in Table 1.

Of course, you will note many elements in my description of the experiment that are not operationally definite. What, for example, are symmetrical chips? What is thorough mixing? How can the experimentalist maintain all other conditions essentially the same? However, this kind of indefiniteness in the operation of drawing at random is much the same as exists in defining any experimental procedure.

The engineer without statistical training is likely to ask what

<sup>3</sup> A more comprehensive discussion of the difference between the mathematical concept of random and the operationally verifiable meaning of random has recently been given elsewhere.

Cf. W. A. Shewhart, *Statistical Method from the Viewpoint of Quality Control*, Graduate School of the Department of Agriculture, Washington, D. C. From the more technical viewpoint there presented, the two random operations here may be treated as one. For our present purpose they can best be considered separately.

See also the discussion of what the mathematician defines as random in the paper by Prof. Wilks, also contributed to this symposium.

<sup>4</sup> Of course there would only be  $N$  different possible orders.

TABLE 1

*Thickness in Arbitrary Units of Inlay on 144 Relay Springs*

|    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|
| 55 | 13 | 45 | 27 | 33 | 59 | 43 | 24 | 43 | 52 | 39 | 62 |
| 72 | 31 | 38 | 36 | 36 | 45 | 26 | 30 | 28 | 42 | 38 | 50 |
| 93 | 37 | 23 | 18 | 34 | 41 | 2  | 28 | 38 | 53 | 49 | 70 |
| 54 | 38 | 58 | 43 | 59 | 47 | 17 | 71 | 28 | 53 | 38 | 60 |
| 30 | 43 | 52 | 10 | 75 | 38 | 28 | 57 | 23 | 42 | 35 | 36 |
| 41 | 38 | 35 | 26 | 59 | 56 | 51 | 37 | 18 | 29 | 35 | 65 |
| 40 | 36 | 50 | 8  | 43 | 39 | 19 | 43 | 39 | 43 | 22 | 50 |
| 45 | 21 | 50 | 53 | 41 | 57 | 42 | 58 | 52 | 17 | 30 | 75 |
| 31 | 30 | 23 | 39 | 57 | 22 | 35 | 33 | 43 | 39 | 24 | 90 |
| 25 | 34 | 68 | 47 | 59 | 43 | 3  | 62 | 31 | 48 | 67 | 96 |
| 30 | 36 | 53 | 48 | 48 | 33 | 27 | 27 | 36 | 58 | 54 | 80 |
| 23 | 41 | 26 | 34 | 42 | 20 | 27 | 26 | 42 | 59 | 73 | 88 |

there is of significance to him about the sequence in Fig. 1a. Well, the answer is that there is something about that sequence that is of very great importance to him. In the first place, most sequences

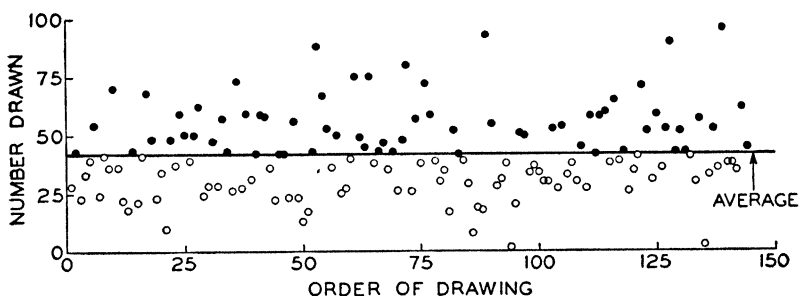


FIG. 1a SEQUENCE OF NUMBERS DRAWN AT RANDOM

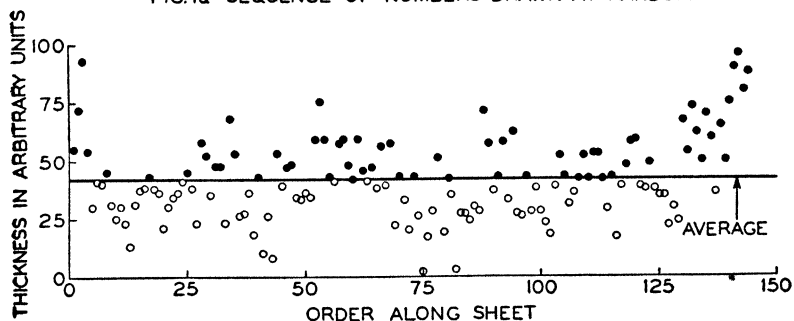


FIG. 1b SEQUENCE OF VALUES OF THICKNESS OF INLAY ON RELAY SPRINGS

obtained by a random operation possess certain characteristics that almost no sequence of results from repetitive engineering

operations is found to possess until after assignable causes have been eliminated through the application of the operation of statistical control. Furthermore, if anyone not an experienced statistician were to try to write down a lot of sequences of 144 different numbers or if he were to try to arrange the 144 numbers in Table 1 in what he would instinctively call random order, most of these sequences would fail to possess the characteristics possessed by almost all of the class of sequences obtained by drawing the 144 numbers again and again at random as described above.<sup>5</sup> Many of my colleagues have told me that their first real feeling for the meaning of a random operation came after they had tried to juggle with their eyes open, as it were, a set of numbers into what they would instinctively call random order only to find that the sequences thus obtained did not possess the characteristics possessed by most of the sequences obtained by drawing from a bowl with their eyes shut.<sup>6</sup> Before we can go further in characterizing the quantitative differences between such sequences, we must consider first the problem of testing a statistical hypothesis and later that of testing the hypothesis that an operation is in a state of statistical control.

*Basic Test of Statistical Hypothesis.* To determine whether any operation such as that of drawing from a bowl, as illustrated by the data in Fig. 1a, gives a sequence with the characteristics of one defined as random by the mathematician, we may choose one or more of the indefinitely large number of criteria that have been or may be established mathematically. If the operation gives a sequence that fails to meet the chosen criteria, the hypothesis is rejected, but if it meets the criteria, the hypothesis is accepted. Needless to say, no experimental test will prove or disprove a

<sup>5</sup> Since there are  $N!$  different possible orders in which  $N$  numbers may be drawn, one might argue that any arrangement whatsoever that one chooses to make of the  $N$  numbers is a random arrangement in that it would be one of the possible  $N!$  orders obtained by a random operation of drawing. The important point to note is that I do not speak of a random number or a random arrangement except in the sense of a number or arrangement given by a random operation. What I am contrasting is the class of arrangements given by the operation of arranging the numbers in what one intuitively may feel is a random manner with the class of arrangements given by the random operation of drawing from a bowl. For a more comprehensive discussion of this point cf. Shewhart, *op. cit.*

<sup>6</sup> The reader may wish to try for himself different orders of the data in Table 1 and see if his trials pass the three criteria considered later in discussion of statistical control theory.

statistical hypothesis, and in fact the test thereof constitutes a *rule of behavior* that must be justified upon the basis of extensive experience showing that in the long run we shall not be too often wrong. In fact, the test of any statistical hypothesis is subject to the following two kinds of errors known as errors of the first and second kinds: (1) sometimes the hypothesis will be rejected even though true and (2) sometimes the hypothesis will be accepted even though false.<sup>7</sup>

Experience reported in the literature from many different sources justifies the conclusion that we may use with confidence the deductive distribution theory of the statistician to predict the distribution of any observed statistic of samples of size  $n$  or of any one of the many characteristics, such as lengths of runs-up and runs-down, of the infinite sequence of numbers that we may expect to get by repeating again and again without limit the operation of drawing a number from a bowl. Certain other operations as, for example, the use of tables of random sampling numbers also give results that have been found by experiment to possess the properties predicted by the mathematical statistician.

Such studies show that a few specific kinds of operation exhibit properties described by the mathematical statistician as random. The ability to randomize a set of numbers or a set of objects by means of some distinguishable physical operation provides the scientist with a powerful technique for making valid predictions, and we shall now see how this can be used by the engineer.

*Four Uses of the Basic Contributions of Classical Statistical Theory.*

(1) Obviously the method of testing a statistical hypothesis can be used for testing the hypothesis that a sample assumed to be random came from an assumed law of chance, or it can be used to test the hypothesis that two samples, both of which are random, came from the same law of chance. Such is the nature of statistical tests of significance of observed differences between two or more

<sup>7</sup>J. Neyman and E. S. Pearson have contributed many important papers on testing statistical hypotheses starting with one in *Biometrika*, July 1928, "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference." Some of their latest contributions are given in *Statistical Research Memoirs*, Vol. 1, 1936, and Vol. 2, 1938, Cambridge University Press, London. Also see "On the Problem of the Most Efficient Tests of Statistical Hypotheses," by J. Neyman and E. S. Pearson, *Philosophical Trans. Royal Society of London*, Series A, Vol. 231, pp. 289-337. It is to be noted that all such tests depend upon the assumption that the sample used in testing is random. Later in testing the hypothesis of statistical control, we start with testing the hypothesis that the sample is random.

samples of data, but it should be noted that these tests are valid only *if* the samples tested are produced by random operations.

(2) If one after another of the operations of the engineer, such as an operation of measurement of some physico-chemical property or that of producing a given kind of product, could be shown to satisfy the hypothesis that it obeyed a law of chance,<sup>8</sup> then each such law and the estimates of the parameters therein could be quite simply obtained by means of well-established statistical procedures. Thereafter an engineer would be justified in using the distribution theory of the statistician in predicting the outcome of any future repetitions of that operation with the same degree of assurance that he has in using statistical distribution theory in predicting the outcome of future drawings from a bowl universe. This constitutes a goal highly to be desired from the viewpoint of design, for then it would be possible to use mathematical distribution theory in establishing the economical overall tolerance limits in terms of those for raw materials and pieceparts.

However, as long ago as 1924, abundant evidence, since substantiated on an even larger scale,<sup>9</sup> was obtained to show that few, if any, operations of the engineer obey laws of chance, even when carried out under presumably the same essential conditions. Two courses of action were open. One was to use available statistical technique as a *curative* measure in the sense of screening the product obtained by repeating a given production operation when this operation does not obey a known law of chance. The other was to develop a statistical technique to be used as a *preventive* measure in the sense of providing a means of detecting and eliminating assignable causes of variability that need not be left to chance.

(3) Next, then, let us see how classical statistical theory provides a means of screening the results already obtained by repeating an operation, as in the production of a product of a given kind. The basis for the technique lies in the empirically established fact that the statistician has justified the use of certain opera-

<sup>8</sup> Needless to say, in practice we can never be *sure* that an operation obeys a law of chance.

<sup>9</sup> Cf. *Economic Control of Quality of Manufactured Product*, W. A. Shewhart, New York, D. Van Nostrand and Company, 1931; *Statistical Method from the Viewpoint of Quality Control*, W. A. Shewhart, Graduate School of the Department of Agriculture, Washington, 1939; and *The Application of Statistical Methods to Industrial Standardization and Quality Control*, E. S. Pearson, British Standards Institution, London, 1935.

tions, like drawing from a bowl or from a set of random numbers, by means of which to randomize a set of objects. With this operation of randomizing, one may break up any quantity of product into a number of lots and establish the most economical plan of sampling these lots to insure both that the producer's risk of having a lot of satisfactory quality rejected will not exceed some previously specified value and that the consumer's risk of accepting a defective lot will not exceed some previously specified value irrespective of the quality of the product in the different lots. These producer and consumer risks correspond to the errors of the two types always involved in testing a statistical hypothesis and were introduced into commercial use within the Bell System as early as 1925 in the development of sampling plans for screening product.<sup>10</sup>

(4) Too much emphasis cannot be laid upon the practical importance of the fact that tests of statistical hypotheses are strictly valid only for random samples. This fact is taken into account in the design of sampling plans for screening, and it must also be taken into account in designing an experiment to test the significance of observed differences in the results obtained by submitting the results obtained by one operation to different subsidiary operations. For example, a given kind of product may be subjected to laboratory or field tests under different conditions. Unless the operation of producing the product obeys a law of chance or, in other words, is in a state of statistical control, it is necessary to randomize the samples submitted to the subsidiary operations in order to obtain a valid test of the significance of the observed differences. If this is not done, observed differences between the results of two or more subsidiary operations may have arisen from assignable differences in the results of the first operation. Application of the operation of randomization is particularly important in the comparison of new designs, new materials or alloys, study of contact phenomena under different conditions, corrosion of materials under different atmospheric conditions, and field trials of equipment, to mention only a few.<sup>11</sup>

<sup>10</sup> Cf. "Method of Sampling Inspection," by H. F. Dodge and H. G. Romig, *The Bell System Technical Journal*, Oct., 1929.

<sup>11</sup> The contributions of R. A. Fisher (see references to follow) are of particular importance in such studies. A simplified treatment of the elementary principles is given in *Field Trials: Their Layout and Statistical Analysis*, by John Wishart, School of Agriculture, Cambridge, England, 1940.



It is important for the engineer to keep in mind when reading all the literature on the randomization of the results of the first operation, that the validity of the tests for significant differences between the effects of different kinds of subsidiary operations rests upon the condition that the latter must be in a state of statistical control although this limitation is not explicitly stated. Hence, to be sure of the validity of tests of significant differences in the effects of subsidiary operations, we must first show that these operations are in a state of statistical control. This caution is necessary because otherwise an engineer may accept the results obtained from small samples at their face value.

By and large, classical theory treats of random fluctuations. It is a theory that tells how phenomena in nature would happen *if* they happened at random as do the results of drawing from a bowl and certain molecular phenomena treated in kinetic theory and statistical mechanics. It tells us how to discover such laws of chance *if* they exist and how to use them for purposes of valid prediction when discovered. It tells us how to make valid test of the hypothesis that a random sample came from a given universe or two random samples came from universes that have certain characteristics in common. It also provides us with a very important experimental operation of drawing at random that can be used in drawing random samples from the results of repetitive operations already made.

It is not, however, a theory designed primarily to tell us whether or not observed phenomena happen at random; or how to attain a state of statistical control (or randomness) of the cause systems underlying the operations of the engineer if these are not already in such a state. Whereas classical statistical theory gives us a very useful curative operation of randomizing results already obtained by a repetitive operation, control theory attempts to give us a very useful preventive operation of modifying the cause system underlying a physical operation or production process until it becomes random in the sense of being in a state of statistical control.

### BASIC CONTRIBUTION OF STATISTICAL CONTROL THEORY

That an ounce of prevention is worth a pound of cure holds for the application of statistics. For example, if a manufacturing process can be made to produce a quality of product distributed

in accord with a law of chance, one can then establish once and for all with the requisite degree of assurance by means of a sufficiently large sample, the probability  $q'$  that the quality of a piece of product will lie within any previously specified tolerance limits  $L_1$  and  $L_2$ . Now, it may be shown rigorously that if one knew this probability  $q'$ , sampling of lots would not tell us anything more about uninspected portions of the lots than we knew before we sampled them.<sup>12</sup> Hence if we could attain this idealized condition, the necessity of sampling would be completely eliminated and there would be no need of applying a screening process. Moreover, if it can be shown under such conditions that it is not possible to modify the manufacturing process further by simply removing assignable causes, then it follows that we have also minimized the percent to be rejected because of failure to meet the tolerance requirements. Let us now see how we may approach this idealized condition with its associated advantages.

*Basic Hypothesis for Statistical Control Theory.* Perhaps, in simplest terms, the fundamental hypothesis is that even though the occurrence of an engineering operation exhibiting a state of statistical control is as rare as the proverbial hen's tooth, it is feasible and often desirable to establish a scientific method of modifying an existing operation until it obeys a law of chance. We shall consider the hypothesis of control in three parts, the first of which is:

*Hypothesis IIa: The maximum attainable degree of validity of prediction that an operation will give a value  $X$  lying within any previously specified tolerance limits is that based upon the prior knowledge that the probability of this event is  $q'$  or more generally upon the prior knowledge of the law of chance underlying the operation.*

This part of the hypothesis is in line with the definite abandonment of the causal laws of classical physics and chemistry in favor of an indeterministic theory that includes the idea of probability in the ultimate laws. It is also in line with current theories of knowledge of the world in that no way has yet been devised for arriving at certain knowledge. So far as the statistical theory of control is concerned, no special attempt is made to justify this part of the fundamental hypothesis. Instead it is simply taken over from modern physics and modern logic. As such it represents the limiting knowledge to which an engineer may hope to attain in giving assurance

<sup>12</sup> For an amplification of this point, see the paper contributed to this symposium by Captain Simon.

that his engineering operations will give results lying within specified tolerance ranges.

*Hypothesis IIb: The maximum degree of attainable control of the cause system underlying any repetitive operation in the physical world is that wherein the system of causes produce effects in accord with a law of probability.*

Such a state of control has been termed a *statistical state* of control. The hypothesis that such a state represents the limit to which one may hope to go in controlling a given operation by finding and removing a few assignable causes of variation was originally suggested by the second law of thermodynamics and its interpretation in kinetic theory and the theory of statistical mechanics. In much the same way that entropy measures the degree of run-downness of a physical system at a given energy level, so the degree of approach to a state of statistical control measures the run-downness of the cause system underlying a given operation. In much the same way that it would take a Maxwell Demon to reverse an otherwise irreversible process, so it may be shown by a study of chance cause systems that it would take the equivalent of this Demon to modify a cause system already in a state of statistical control without changing, as it were, the whole operation and hence the whole system of causes.

On many occasions, I have heard an engineer say on reaching a state of statistical control of some operation that, without changing the kind of operation, he was going to decrease the range of variation still further by simply asking the men performing the operation to take greater pains to reduce the variability, but I have not yet witnessed one case where such effort succeeded. Hence the engineer is pretty safe in taking the state of statistical control as a limit to which he may go in reducing the tolerance range for that particular operation.

*Hypothesis IIc: It is assumed that some criterion or criteria may be found and methods developed for their application to the numbers obtained in a sequence of repetitions of any operation such that whenever a failure to meet the criterion or criteria is observed, an assignable cause of variability in the results given by the operation may be discovered and removed from the operation. It is further assumed that, by the removal of a comparatively small number of causes, a state of statistical control is approached where the results of repetitions of the operation behave in accord with a law of chance.*

The development of an operation of statistical control and its use in justifying hypothesis IIc is an empirical contribution of mass production because only in such a process would it be economically feasible to make the wide range of trials of experimental techniques for the purpose of finding one that works satisfactorily.

Let us now consider the requirements in the way of experimental data for attaining control.

*Basic Experimental Data for Attaining Control.* Let us consider the second or experimental step in the scientific method of using control theory. The crucial difference between the experimental technique in control work and that in applying the classical theory of statistics to provide a screen is that in control work we pay attention to the condition  $C_i$  under which an operation gives a value  $X_i$ , whereas in the screening operation we ignore this condition; in control work, the experimentalist must keep his eyes very much open, but in screening product he must, as it were, keep them blindfolded; and in control work, interest centers in controlling the product not yet made through modification of the underlying cause system, but in the screening process, interest centers in the product already made.

To symbolize this situation let us rewrite sequence (1) and attach to each value  $X_i$  its associated symbol for condition  $C_i$ . Also let us divide the sequence into two parts representing that already observed and that observable in the future. Then we have

$$\begin{array}{ccccccc|cccc}
 X_1, & X_2, & \dots, & X_i, & \dots, & X_j, & \dots, & X_n, & & X_{n+1}, & \dots, & X_{n+k}, & \dots \\
 | & | & & | & & | & & | & & | & & | & \\
 C_1 & C_2 & & C_i & & C_j & & C_n & & C_{n+1} & & C_{n+k} & \\
 | & | & & | & & | & & | & & | & & | & \\
 \text{Past} & & & & & & & \text{Present} & & & & \text{Future} & 
 \end{array} \quad (2)$$

The screening process applies to the  $X$ 's of the past and eliminates the information contained in the  $C$ 's through the operation of drawing samples at random. It also completely ignores the results of future repetitions. In contrast, the control process focuses attention on both the  $X$ 's and the  $C$ 's of the past in the hope of detecting and removing assignable causes in the  $C$ 's of the past so that these causes will not enter into future repetitions of the operation. Of course, in practice the screening and control processes may be carried on simultaneously and the data obtained in the screening process may also be used for the purpose of detecting lack of control. For example, inspection samples of lots taken in the order of

their production have long been used in some industries in applying the operation of statistical control shortly to be described.

Just as the batter must keep his eye on the ball first to make a hit and afterwards to determine what to do, so does the experimentalist have to keep his eye on the condition underlying the performance of any operation first to be able to apply this information most effectively in testing the hypothesis that he can do something to change the condition, and afterwards to determine what to do if the test is positive. By keeping his eye on the condition associated with each of  $n$  repetitions of an operation, the engineer may distinguish the following three situations:

(1) In the absence of any a priori reason for distinguishing between any two conditions, he may judge them to be essentially the same as may be symbolized by the equivalence

$$C_1 \approx C_2. \quad (3)$$

Operations of drawing with replacement from a bowl illustrate a situation where the experimentalist is practically forced to conclude that the conditions are essentially the same from drawing to drawing. However, even under such circumstances, it is possible to arrange the results obtained by repeating an operation again and again in the same order as the operation was repeated.

Now, if the repetitions of the operation do not obey a law of chance or, in other words, do not arise under a state of statistical control even though the experimentalist considers them to be essentially the same, as symbolized by (3), the sequence of values of  $X$ , when taken in the order that they were observed, is not likely to pass the criteria of randomness established by the statistician. In fact, the order of observation in such instances provides the only quantitative basis for testing the hypothesis that the observations came from a state of statistical control, and in practice it is usually found that this order indicates lack of randomness. Hence the statistical control engineer should insist that the record of the order of repetitions be preserved even though the experimentalist judges the conditions to be essentially the same.

(2) The engineer may have a priori reasons for believing that the conditions do not remain essentially the same, as can be symbolized by

$$C_1 \neq C_2. \quad (4)$$

For example, he may surmise that there are erratic effects or possibly trends superimposed upon the effects of a law of chance. In

such a situation, the experimentalist may be in a position to suggest certain ways of ordering the results of  $n$  repetitions of an operation, independently of the magnitudes of the associated values of  $X$ , and solely upon his knowledge of the conditions under which the observations were taken, in order to reveal, if present, trends or erratic effects that he considers likely to exist.

For example, in some recent work, interest-centered in attaining a state of statistical control of the variation in thickness of a rolled inlay of contact metal on a particular kind of relay spring. These springs were cut from a long strip and there were a priori reasons for expecting both trends and erratic effects in the thickness along the strip. Hence the springs were numbered in the order that they were cut from the original strip so that the measurements of thickness on the individual springs could be arranged in the order of their original position in the strip. These observations have already been given in Table 1. The order in which the observations were arranged along the strip is that of the table, beginning at the left and reading down the columns. The set of 144 ordered measurements is given in Fig. 1b and shortly we shall see how useful this experimental order is in giving clues to the presence of assignable causes.<sup>13</sup> Hence, because of the importance of order, the statistical control engineer must insist that the experimentalist suggest orders in which to arrange the results of a series of  $n$  repetitions when he does not have any a priori reason for believing that any two repetitions have been carried out under the same essential conditions.

(3) The engineer may have a priori reasons for believing that the conditions may be divided into rational groups such that within each group the conditions are essentially the same but such that the conditions for each group are not essentially the same as those for any other group. In this situation, the engineer must insist that the experimentalist indicate the groupings and also the observed order within each group so that this order as well as the differences between the groups may be tested for indications of assignable causes.

Thus far we have considered three kinds of information that the statistical control engineer needs to know about the conditions associated with the repetitions of an operation so as to make the quantitative results useful in testing the hypothesis that the opera-

<sup>13</sup> I am indebted to my colleague, Mr. E. B. Ferrell, for permission to use these data and for many helpful suggestions and criticisms about the scope of applications of a statistically scientific method as treated in this paper.

tion is in a state of statistical control. However, the contribution of the statistician does not necessarily end with the attainment of such a state, because even then the variance and hence the economic tolerance range may exceed the desired value.

Accordingly, it may be desirable to change the whole operation or some part thereof. For example, even though the variations in the thickness of the inlay of contact material shown in Fig. 1b were found to have risen from a state of statistical control, the variance might still be so large that many units of product would have to be rejected in assembly. Under such conditions, there would be no use trying to find and remove assignable causes of variability because these would have presumably been removed. If because of too large variance, the whole operation of producing the inlay is to be changed, the job is primarily one for the engineer, but if only some of the component parts of the operation such as the kind of inlay or fuse metal used, the rolling process, method of inserting the inlay, or heat treatment of the material are to be changed in order to reduce the overall variance, the statistician may be of great service in showing how to obtain at minimum cost the data necessary for analyzing the total variance of the operations into the component parts associated with the component parts of the operation of production.<sup>14</sup>

If some one of the component variances is significantly larger than the others, then it may be most economical to change, if possible, the corresponding component of the operation of production. In any case, the technique of analysis of variance makes it possible to estimate the maximum reduction in total variance that may be expected as a result of changing any component of an operation in a state of statistical control even though no substitute for this component is known at the time. It is very important, however, for the engineer to note that, in accord with hypothesis IIb, the problem of reducing the variance in the results of an operation already in a state of statistical control by changing one or more of the component operations is a fundamentally different<sup>15</sup> problem from that of eliminating assignable causes

<sup>14</sup> The principles underlying the design of efficient experiments for analyzing variance have been set forth by R. A. Fisher and his co-workers. See, for example, *Statistical Methods for Research Workers*, 7th ed., 1938, and particularly *The Design of Experiments*, 2nd ed., 1937, Oliver and Boyd, London. Also see *The Methods of Statistics*, by L. H. C. Tippett, 2nd ed., Williams and Norgate, London, 1937, for applications in the cotton industry.

<sup>15</sup> This significant practical difference has, in general, been overlooked by many students of statistical theory.

without changing the component part of the operation such as kind of inlay used, the rolling process, and the like in the example considered above.

When assignable causes are present, the engineer may reasonably expect to find and eliminate these without changing the production operation or any of its component parts. For example, if assignable causes of variation are present in the process of producing the thickness of inlay on relay springs as shown in Fig. 1b, the engineer may find that they could be removed by further adjustment of the controls of component operations of rolling, heat treating, and the like. When assignable causes are shown not to be present, and it is found that one of the component operations, say the rolling operation, contributes a large share of the variance, it will usually be most economical to secure a reduction in variance by changing this component operation, perhaps by using another type of rolling mill.

Enough has been said, I hope, to indicate some of the things that an experimentalist must do in taking and recording data if he is to obtain data that can be used efficiently in testing the hypothesis of statistical control and in analyzing the total variance under controlled conditions into its component parts or, in other words, if he is to make progress in the removal of assignable causes of variability that need not be left to chance in almost every field of science and engineering, thereby extending the potential usefulness of raw and fabricated materials.

Even in the face of this situation, however, it is still customary engineering practice to neglect the importance of order and instead group together into a frequency distribution all data whether  $C_i \approx C_j$  or  $C_i \neq C_j$ , and irrespective of whether or not subgroups of conditions might be suggested. Then most likely engineers will use the average of this distribution and in addition they may possibly plot the distribution as a frequency curve or ogive.

For example, in making field studies on new materials or designs, studies of the electrical properties of contact materials, or the corrosion of materials, the engineer's faith is likely to be pinned on the averages of large numbers of observations whereas such a procedure is almost sure to mask the very differences that he is looking for. In fact, we shall shortly see that where there are assignable causes present, as soon as we group data together we are almost certain to destroy thereby all clues to the presence of



these causes, the effects of which must be eliminated before we can make valid comparisons of materials, designs, and the like.

*Basic Test of Statistical Control Hypothesis.* The test of the hypothesis that a technique can be established for approaching a state of statistical control *must be empirical*. One sets up criteria of control and then searches for and eliminates, if possible, the causes producing the deviations that fall outside the limits fixed by the criteria. If such search reveals the presence of assignable causes and if, as these are removed, one approaches a state where the criteria of control are satisfied, we accept this evidence as an empirical justification of the hypothesis.

A satisfactory technique for attaining a state of statistical control has been described in considerable detail elsewhere<sup>16</sup> and consists of the following five essential steps that are referred to in control theory as the *operation of statistical control*:

(1) Specify in a general way how an observed sequence of  $n$  data is to be examined for clues to the existence of assignable causes of variability. For example, it is essential that the order in an observed sequence always be tested for randomness whether  $C_i \approx C_j$  or  $C_i \neq C_j$ .

(2) Specify how the original data are to be taken and how they are to be broken up into subsamples upon the basis of human judgments about whether the conditions under which the data were taken were essentially the same or not.

(3) Specify the criterion of control that is to be used and indicate what statistics are to be computed for each subsample and how these are to be used in computing action or control limits for each statistic for which the control criterion is to be constructed. Three of the conditions that such criteria should satisfy are as follows: the limits in the criteria should be as nearly independent as possible of the functional form of the law of chance when the state of statistical control is attained; the criteria should in so far as possible minimize the error of accepting the hypothesis when false and should keep the error of rejecting the hypothesis when true less than some prescribed value fixed by economic considerations; and the criteria should indicate as closely as possible the condition under which the assignable causes enter the operation and as much as possible about the nature of these causes.

(4) Specify the action that is to be taken when an observed statistic falls outside its control limits. The general action required

<sup>16</sup> *Statistical Method from the Viewpoint of Quality Control, loc. cit.*

is to look for assignable causes whenever the criteria are not satisfied.

(5) Specify the quantity of data that must be available and found to satisfy the criterion of control before the engineer is to act as though he had attained a state of statistical control.

To illustrate the role played by statistical criteria in the process of detecting and removing assignable causes of variation and to show how important it is not to group data together so as to destroy all possibility of ordering them in terms of the conditions under which they were taken, let us see what happens when certain of these criteria are applied to the two sequences of Fig. 1. Will the criteria tell us anything about these two sequences that we cannot see by just looking at the sequences themselves? To comprehend the significance of order, it should be kept in mind that the data in Fig. 1a are the same as in Fig. 1b except that in Fig. 1a they appear in the order drawn from a bowl. Anything that we can find out about the presence of assignable causes by studying the order in Fig. 1b is thus lost just as soon as this original order is destroyed by grouping the data together. Will the criteria of control indicate the presence of assignable causes of variation in both Fig. 1a and 1b? Of course, we know that if they give an indication of such causes for the drawings from a bowl (Fig. 1a), this will likely constitute a false lead because experience shows that we should not expect to find any assignable causes in the operation of drawing from a bowl, particularly when carried out by an experienced observer.

As a starting point, let us apply the control chart criterion based upon averages of successive fours.<sup>17</sup> Fig. 2 shows what happens: No false lead is given for the drawings from the bowl, Fig. 2a, but the presence of assignable causes of variation in thickness is indicated, Fig. 2b.

<sup>17</sup> For a detailed statement of the method of constructing a control chart, see *Economic Control of Quality of Manufactured Product*, loc. cit., pp. 309-313. It should be kept in mind that, for reasons that I have given in the literature, it is desirable to use small subgroups as is here done. See, for example, Chapter 1 of *Statistical Method from the Viewpoint of Quality Control*, loc. cit. Of course, the practical man wants to know what is a small sample, and to him it may be of interest to know that in much of my own work, particularly in laboratory research, I have found it desirable to use where possible subgroups of size four. One reason for not using a larger sample is that, as we shall soon see, runs of seven or more are very unlikely for a statistically controlled process. Hence, if we are to be sure that at least one subgroup will be completely within a run of seven, we should not use a subgroup larger than four.

Next let us see if other criteria may be found that will tell us something about the nature of assignable causes to be expected. For example, do you detect any evidence of causes producing

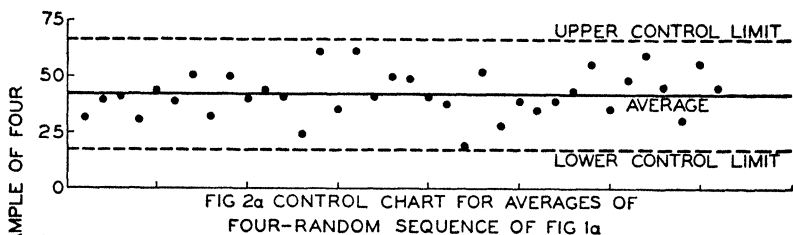


FIG 2a CONTROL CHART FOR AVERAGES OF  
FOUR-RANDOM SEQUENCE OF FIG 1a

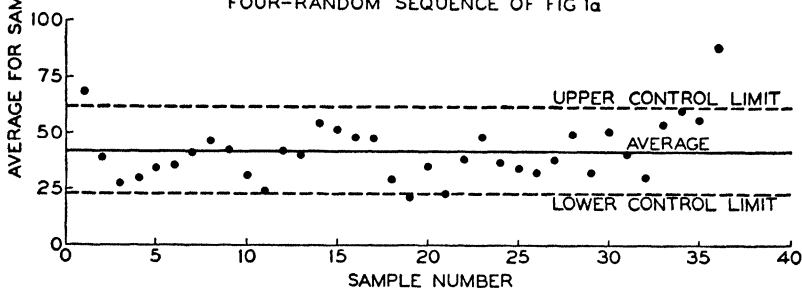


FIG 2b CONTROL CHART FOR AVERAGES OF FOUR  
SEQUENCE OF THICKNESS OF FIG 1b

trends in the data of Fig. 1b? Is there a downward trend at the left and an upward one at the right? Or would you expect to find that the assignable causes are of the type that produce discontinuous and erratic shifts in the expected values? Is there any evidence for believing that both kinds of causes are present? It should be kept in mind that we seek answers to these questions so that we may be better able to detect and remove the assignable causes.

Recent studies covering a broad field of research problems have shown that two very simple criteria may be used successfully in helping to distinguish between causes producing trends and those producing discontinuous and erratic effects. A complete report on the application of such criteria cannot be given here and it must suffice to indicate in broad outline the simple nature of these criteria and how they have been found to work. Basically, what is done is to note two kinds of runs in any sequence: runs up and down and runs of numbers above and below the average of the sequence where runs up and down are defined as follows. In any sequence, certain numbers are greater than either of their im-

mediate neighbors and form, as it were, maxima in the sequence. In a similar way, we have minima. A run-down is the interval between a maximal number and the next succeeding minimal one, and the length of run-down has been defined in the literature as the number of numbers involved, including the maximal and minimal numbers. Corresponding definitions hold for runs-up.

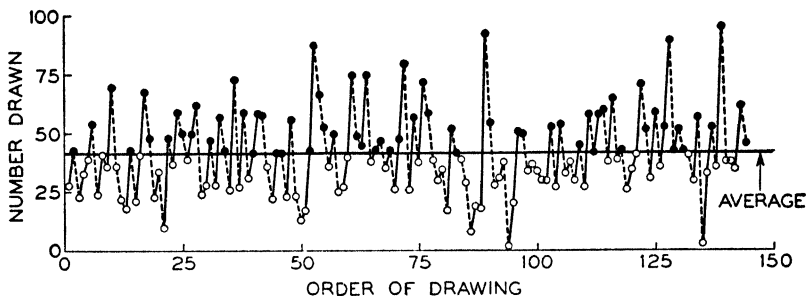


FIG 3a SEQUENCE OF NUMBERS DRAWN AT RANDOM

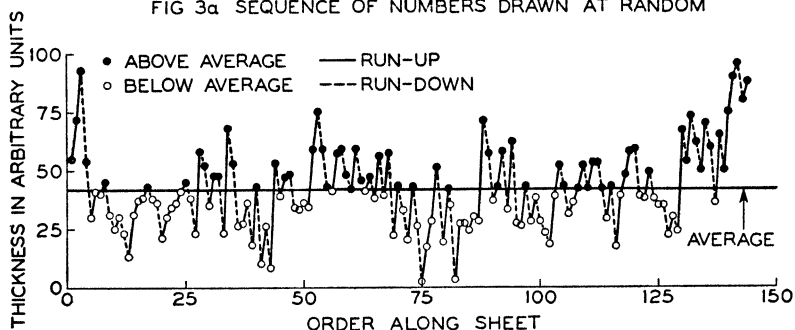


FIG 3b SEQUENCE OF VALUES OF THICKNESS OF INLAY ON RELAY SPRINGS

In Fig. 3 where the data of Fig. 1 are replotted, runs-up are shown by solid lines connecting observed values, and runs-down by dotted lines. Also in Fig. 3 the points above the average of the sequence are filled in and those below are not.

W. O. Kermack and A. G. McKendrick<sup>18</sup> have recently given

<sup>18</sup> "Tests for Randomness in a Series of Numerical Observations," *Proc. of Roy. Soc. of Edinburgh*, Vol. LVII, pp. 228-240, 1937. If the total number of observed runs up and down be  $N$ , these authors show that the expected number of lengths 2, 3, 4 . . . , are  $0.6250000N$ ;  $0.2750000N$ ;  $0.0791667N$ ;  $0.0172619N$ ;  $0.0030506N$ ;  $0.0004547N$  and  $0.0000587N$ . See also R. A. Fisher's note, "On the Random Sequence," in the *Quarterly Journal of the Royal Meteorological Society*, July, 1926, p. 250. Of course, the probabilities here given apply to a frequency distribution of runs in an infinite sequence. By making use of results given by Kermack and McKendrick in a paper, "Tests for Randomness in a Series of Numerical Observa-

the probability of a run of any length  $l$  occurring in an observed set of  $N$  runs up and down from any infinite random sequence independent of the law of distribution of the variable. Likewise, W. G. Cochran<sup>19</sup> has recently given a formula for deciding whether sequences of similar meteorological events, such as runs of consecutive wet months, may be expected under a state of statistical control. If the result of an operation must be either  $E_1$  or  $E_2$ , and if we know the probability  $p$  that an operation will give the event  $E_1$ , and the probability  $(1 - p) = q$  that the operation will give the event  $E_2$ , then Cochran's work gives the expected number of runs of length  $r$  (counting runs of both  $E_1$  and  $E_2$ ) in any sequence of  $m$  repetitions of the operation. If now we call  $\bar{X}$  the average of a sequence of  $m$  observed numbers and if we call event  $E_1$  the occurrence of a number in the sequence greater than  $\bar{X}$  and event  $E_2$  the occurrence of a number less than  $\bar{X}$ , we may use Cochran's theory to compare the observed number of runs of any length  $r$  with the corresponding theoretical number to be expected upon the *assumption* that the observed fraction of the numbers greater than the average  $\bar{X}$  is equal to the probability  $p$  that the underlying cause system will give a value of  $X$  greater than the number  $\bar{X}$ , considered simply as a number and not necessarily as the average.<sup>20</sup>

Now, if the assignable causes simply produce erratic shifts in the expected values, it is not likely that the distribution of lengths of runs-up and of runs-down will be much disturbed, but such shifts will tend to give extra long runs of numbers above and below average. However, if the assignable causes produce trends with

---

tions," *Proceedings of the Royal Society of Edinburgh*, Vol. LVII, Part III, No. 17, pages 352-356, it may be shown that there will not be appreciable differences in this frequency distribution so far as we are here concerned with its use even though we are dealing with an observed distribution of runs from a finite sequence of 144 numbers.

<sup>19</sup> Cf. "An Extension of Gold's Method of Examining the Apparent Persistence of One Type of Weather," in *Quarterly Journal of the Royal Meteorological Society*, Vol. LXIV, pp. 631-634, 1938. The number of runs of length  $r$  of the events  $E_1$  and  $E_2$  out of  $m$  trials is

$f_{m,r} = 2(p^r q + p q^r) + (m - r - 1)(p^r q^2 + p^2 q^r)$  when  $1 \leq r \leq (m - 1)$   
where  $p + q = 1$ .

<sup>20</sup> Of course, one might arbitrarily choose any value  $X_1$  and any value  $p$  and use Cochran's theory to compare the observed number of runs of length  $r$  both above and below  $X_1$  with the corresponding theoretical number to be expected upon the assumption that the probability of occurrence of a number greater than  $X_1$  is  $p$ .

slopes large in comparison with the fluctuations produced by the superimposed causes acting at random, the distribution of lengths of runs-up and runs-down will be disturbed. Likewise, the distribution of runs of numbers above and below the average of the sequence may be somewhat modified.

Of course, we should expect close agreement between theory and experiment for the lengths of runs of both kinds in Fig. 3a, but since the control chart Fig. 2b gave evidence of the presence of assignable causes, we may expect discrepancies between theory and experiment for the lengths of runs in Fig. 3b. The data for such comparisons are given in Table 2. As expected, there is close

TABLE 2

| Length<br>of<br>runs | <i>Drawings from bowl</i>       |                         |                                 |                                 | <i>Measurements of inlay thickness</i> |                         |                                 |                                 |
|----------------------|---------------------------------|-------------------------|---------------------------------|---------------------------------|--|-------------------------|---------------------------------|---------------------------------|
|                      | Runs above and<br>below average |                         | Runs up and<br>down             |                                 | Runs above and<br>below average        |                         | Runs up and<br>down             |                                 |
|                      | Ob-<br>served<br>number         | Ex-<br>pected<br>number | Ob-<br>served<br>fre-<br>quency | Ex-<br>pected<br>fre-<br>quency | Ob-<br>served<br>number                | Ex-<br>pected<br>number | Ob-<br>served<br>fre-<br>quency | Ex-<br>pected<br>fre-<br>quency |
| 1                    | 42                              | 36                      |                                 |                                 | 29                                     | 36                      |                                 |                                 |
| 2                    | 18                              | 18                      | 65                              | 61                              | 13                                     | 18                      | 53                              | 57                              |
| 3                    | 10                              | 9                       | 22                              | 27                              | 3                                      | 9                       | 28                              | 25                              |
| 4                    | 4                               | 4                       | 10                              | 8                               | 5                                      | 4                       | 7                               | 7                               |
| 5                    | 4                               | 2                       | 1                               | 2                               | 0                                      | 2                       | 2                               | 2                               |
| 6                    | 0                               | 1                       | 0                               | 0                               | 4                                      | 1                       | 1                               | 0                               |
| 7                    | 0                               | 1                       | 0                               | 0                               | 4                                      | 1                       | 0                               | 0                               |
| 8                    | 0                               | 0                       | 0                               | 0                               | 1                                      | 0                       | 0                               | 0                               |

agreement for the drawings from the bowl. For the sequence of thickness measurements, there is excellent agreement between theory and experiment for the distribution of lengths of runs-up and runs-down but not for lengths of runs above and below average.<sup>21</sup> These results constitute good evidence for believing that the assignable causes underlying the measurements of thickness in Fig. 3b do not produce trends such as might be produced by lack of symmetry in the rolls but simply produce discontinuous and erratic shifts in expected values of thickness such as might be produced by slippage at the cleavage planes of the crystals in the inlay.

Not only does the application of these criteria to the observed

<sup>21</sup> Of course, the total number of expected runs above and below average will not usually be the same as the corresponding number of observed runs.

order indicate the presence of assignable causes and give a clue to their nature, but it also indicates when the assignable causes enter. For example, we see from the table that runs above and below the average of length greater than 7 are very unlikely,<sup>22</sup> and all we need to do is to note where these runs occur, keeping in mind that such indications are subject to errors of the first and second kind as described above.

Since all of this information about assignable causes is lost just as soon as we group the data together in the form of a frequency distribution, as is so often done by the engineer, we see how very important it is to pay attention to order in taking and recording data. For example, years of experience have shown that Nature almost never gives us a sequential order that passes even the control chart criterion and is almost certain not to give us one that passes all three criteria and that the failure to meet such criteria almost always is traceable to an assignable cause. Yet, if we were to take these same data without reference to observed order as might be done by shaking them up, drawing them at random, and then applying the three criteria, almost all of them would slip through the net, thereby failing to indicate the presence of assignable causes that must be detected and removed in order to attain the advantages of a state of statistical control.

Since, as noted above, experience has shown that one can almost always find and remove assignable causes when indicated by such criteria and thereby approach a state of statistical control beyond which it is not possible to go except by some fundamental change in the operation itself, we have good grounds for accepting the basic hypothesis of statistical control theory as stated at the beginning of this section.

## SUMMARY OF POTENTIAL CONTRIBUTIONS OF STATISTICS TO THE SCIENCE OF ENGINEERING

The basic contribution of statistics to the science of engineering is an *improved scientific method to fit the world of probability in which we live*. Classical theory contributes the hypothesis of a repetitive operation obeying a law of chance, the knowledge of which combined with the knowledge of formal mathematical distribution theory enables an engineer to make valid predictions of the out-

<sup>22</sup>The expected number is actually 0.5682 instead of the rounded value of unity given in Table 2.

come of future repetitions of the operation. Statistical control theory contributes the hypothesis that it is humanly possible to remove assignable causes of variability in the repetitive operations of the engineer until such operations approach a state of maximum control and obey laws of chance, a knowledge of which provides maximum assurance that the results of repeating an operation will fall within previously specified tolerance limits. To test these two hypotheses, statistical theory provides the necessary experimental techniques outstanding among which are (a) the operation of randomization and (b) the operation of statistical control.

Broadly speaking, statistical theory treats of repetitive operations and provides the engineer with a method of regulating such operations to his best interest. Fundamentally, the engineer's job is to devise operations that, if carried out, will give results within previously specified tolerance limits. Sometimes the operation, like that of building a bridge, is to be carried out only once or at most a few times, and sometimes the operation, like that of mass production, is to be carried out an indefinitely large number of times. Inherently, all such operations are potentially repetitive, and in this sense, differ only in the number of repetitions carried out. It has long been recognized that one of the most revolutionary principles ever introduced into manufacturing was that of interchangeability dating back at least to Eli Whitney in 1798. The introduction of that principle prompted the engineer to consider the advantages of introducing repetitive operations into production processes, and the contribution of statistics to engineering may be thought of as a means of maximizing the advantages to be attained by interchangeability.

The basic contributions of statistics to scientific method make possible the attainment of the following objectives that are not otherwise attainable and that are of interest to all of us:

(1) Even before a repetitive operation has reached a state of statistical control, the application of statistical theory makes possible the establishment of sampling plans that will screen at minimum cost the output of such an operation so as to meet previously specified tolerance requirements and previously specified producer and consumer risks.

(2) The use of statistical theory provides efficient experimental techniques based upon the operation of randomizing the results obtained from one operation or production process that is not in



a state of statistical control before submitting them to two or more subsidiary operations or treatments for the purpose of comparing the effects of these subsidiary operations, as in the case of field tests and the like. Such a procedure minimizes the chance of concluding that observed differences in the effects of the subsidiary operations are significant when in fact they came about because of assignable differences in the results of the first operation.

(3) The operation of statistical control provides an experimental technique for minimizing tolerance ranges and maximizing the assurance that the product turned out by a given process will meet its tolerance requirements. Such an operation makes possible the most efficient use of limited quantities of raw material and provides the maximum degree of refinement attainable by any production process. Preliminary studies indicate that the operation of statistical control also provides a useful technique for eliminating assignable causes of variability in certain kinds of human effort as, for example, typing and other forms of transcription. Both strategically and commercially, industrial groups and even nations often need every increment of efficiency in the use of limited quantities of raw materials and human effort that can be provided through the application of the operation of statistical control. Likewise they often need maximum refinement in quality through elimination of assignable causes, not only in pursuit of the arts of peace but also in time of war. As one example, the attainment of maximum homogeneity and hence minimum tolerance ranges in the properties of raw and fabricated materials may extend the potential carrying capacities of ships in the air and on the sea. Needless to say both the engineer and the consumer of the engineer's or manufacturer's products stand to gain through the increased assurance that the products will be found to meet their tolerance requirements.

(4) The operation of statistical control provides a technique for modifying and coördinating the three fundamental steps in the process of mass production, namely, specification, manufacturing, and inspection, so that the maximum number of pieces of product having a quality within specified tolerance limits can be turned out at given cost. It does this by showing how to minimize the cost of inspection and the cost of rejection.

In conclusion it may be said that statistical theory plus mass production provides a means of maximizing our physical comforts in time of peace and our strategic factors in time of war.

UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

Contribution of Statistics to the Development  
and Use of Purchasing Specifications  
and Standards of Quality

*By*

LESLIE E. SIMON\*

INTRODUCTION

ALL of you have heard of the blind man in a pitch-dark room trying to find a black cat that was not there. I believe that a man would be in much the same position if he tried to write a specification without statistics. I say this, knowing full well that quite workable specifications are written by men who have no knowledge whatever of formal statistics. However, these men are none the less users of statistics. Without attempting to credit all successful achievement to statistics, it is evident that by keen native judgment and observation, these men have indirectly attained satisfactory specifications. But the problem of dealing with a variable product is fundamentally statistical, and these men could have achieved their desired end much more economically, swiftly, and unerringly with the proper statistical tools. The distinction between the use and non-use of statistics in this situation is perhaps similar to that between the two hunters, one of whom is armed with a spear, the other with a rifle. Both may succeed in eating, but one must strive for a precarious existence, while the other may live in relative comfort. So in writing good specifications: they may be eventually reached by the brute-force method of trial and error and hard common sense, but can be attained with greater ease and certainty through the use of the scientific aids offered by statistical theory. It is of the contribution made by this powerful auxiliary that I wish to speak.

Several years ago I was given what I believe to be an unusual opportunity to observe the functioning of purchasing specifica-

\* Captain, Ordnance Department, United States Army; Executive Officer, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland.

tions, their relationship to standards of quality, and the power of the statistical technique in ascertaining and defining quality with swiftness, clarity, and validity.

Since this experience is of a military character I shall, for obvious reasons, confine myself to broad scientific principles which are of general application both to the Ordnance problems and to many problems in industry. I may, however, define the experience by stating the general circumstances.

It is necessary for the Army to maintain stocks of ammunition. Unfortunately ammunition is subject to deterioration in storage. It therefore follows that, after a period of years, investigations must be conducted to determine the nature and extent of the deterioration, and lots of ammunition must be assigned to various grades, with a view to their ultimate replacement when their quality becomes too poor.

In order to detect deterioration, one must know what the quality was originally. It is but natural, therefore, to look to the procurement specification as at least an indication of the desired initial quality. Since there is frequently some doubt about the meaning or efficiency of the procurement specification, one supplements this knowledge by a critical examination of all the acceptance tests, subsequent tests, and recorded service use which pertain to the ammunition. This procedure gives a keen insight into both the level of quality which was current at the time, and the efficiency of the procurement specification as an instrument for obtaining a product of the apparently intended level of quality. Finally, tests must be performed to determine (at least within limits) the existing level of quality after deterioration, if any, has taken place. Since the function testing of ammunition is both destructive and costly, it is highly desirable that all possible information be extracted from such expensive test data. Statistical methods are therefore used in the analysis of test data. They provide reliable protection against the frailty of human judgment in estimating the quality of lots of ammunition from the results of tests of samples taken from those lots, and are a cogent factor in making judgments comparable when made from time to time and by person to person. This procedure throws much light on the operation of the statistical method and the results which might have been obtained, had these methods been in use at the time the ammunition was originally accepted. At least some of these principles have now been applied to current procurement specifications. I should like to take

this opportunity to express my gratitude to Dr. Walter A. Shewhart for his generous help and patient advice while I was devising the system of surveillance of ammunition which has now been in use for over two years.

## THE OLDER METHOD OF PROCUREMENT

Some of the contributions of statistics to the development of specifications and standards of quality can best be shown by first analyzing the conditions which exist in the absence of the statistical method.

The attainment of one's professional end under these circumstances is sometimes highly imperfect, slow, and costly. When it becomes necessary to procure a product in quantity there is generally a considerable degree of incertitude both as to the level of quality which is necessary and sufficient for the purpose intended and as to the level which is practically attainable in manufacture. These levels, if clearly defined, would surely form the basis of a sound and economical standard of quality. It appears that the management frequently avoids the tedious problem of determining an economic standard of quality, of evaluating the statistical constants associated therewith, and of devising an inspection procedure to insure to an appropriate degree the meeting of the standard by what appears to be a clever and practical stratagem. They merely decide not to bother about either the standard or the statistics (perhaps on the grounds that it is impractical or at least troublesome), and instead write a good, stiff acceptance specification with a view to getting the very best obtainable. I think that sly old Aesop could have written a good fable about this attitude. It carries grave penalties.

Of course, an acceptance specification calls for inspection. Inspection is always costly, and therefore some sampling scheme is generally adopted in the interest of economy. In the case of ammunition, sampling must suffice, since a functioning test of the entire product results in its total destruction. The specification therefore generally reduces to one or more requirements of the form, "a test sample of  $n$  shall be selected at random from each lot, batch, or consignment of the product, and not more than  $c$  shall fail." The  $n$  used in most of these specifications is very small. Many of you have or will come in contact with this type of specification. I wish to show that this popular and well-known type of specification

offers almost no guarantee of the quality of product which will be accepted thereunder, and that the quality of product accepted under such conditions is practically that which happens to be offered for inspection.

Suppose, for example, that a manufacturer submits a lot for acceptance of which the fraction  $p$  are good and the fraction  $q$  are failures, or more specifically  $q$  is the fraction defective. It is easy to show by the binomial distribution that such a lot has a definite probability,  $P_b$ , of passing the above specification where

$$P_b = \sum_{k=0}^c \frac{n!}{(n-k)!k!} p^{n-k} q^k$$

It is thus evident that if manufacturer A makes a very good, but not quite perfect product, some of his lots will be rejected; and if manufacturer B presents only very bad lots, some of his lots will be accepted. In general, if a large number of lots, all having the fraction defective,  $q$ , are presented for acceptance, then  $100 P_b\%$  of the lots will be accepted and  $100 (1 - P_b)\%$  of the lots will be rejected. The most discouraging factor, however, inheres not in the rejection of some good lots or in the acceptance of some bad lots but in the fact that the accepted lots in this case are no better than the rejected lots. Suppose that  $k$  manufacturers produce lots each of fraction effective  $p_1, p_2, \dots, p_k$ , respectively, where  $p_1$  to  $p_k$  represents a limited range which in practice may be of the order of 0.99 to 0.85. Each manufacturer will have his lots rejected with more or less frequency, but the accepted lots of a manufacturer will be no better than the rejected lots. Whereas no precise statement can be made as to the proportion of accepted product at any given quality level without a knowledge of the number of lots presented at each quality level, it is nevertheless obvious that the average of the accepted quality will differ little from the average of the presented quality, and that the limits are the same, even though one require that there shall be no defectives in a small sample of  $n$ . For the case of zero defectives in a sample of ten, and equal frequency of all kinds of lots from 0.99 to 0.85, the average quality levels are 0.92 (presented) and 0.94 (accepted). For a specification of not more than one defective in ten, they become 0.92 and 0.93. It is therefore clearly evident that the quality of product accepted under the old type of specification is practically that which happens to be offered for inspection. It also appears

that the only way to prevent the acceptance of poor quality consists in not allowing it to be presented for inspection. More will be said of this subsequently.

This unhappy situation cannot be corrected by any simple remedy, for the same law applies whatever the value of  $c$ . Suppose that one reduces  $c$  to zero, in an effort to improve quality. What is the result? A plot of the binomials, which in surveillance is called the operating characteristic of the specification, readily shows that the frequency of rejections of lots of poor quality is somewhat increased; so is the frequency of rejections of lots just short of perfection. The penalties fall on both the just and the unjust as one is merely subjected more mercilessly to the rigors of an unjust chance, and the other merely permitted to profit less from its vagaries. The unsatisfactory condition remains unaltered, and additional grounds for producer dissatisfaction may have been introduced. Quality becomes a function of margin of profit: the larger the margin of profit, the greater the number of rejections the manufacturer can afford; and, consequently, if it costs more to produce good quality than poor quality, the more he can consciously lower his level of quality in the interest of making maximum profit. On the other hand, one does not get what he does not pay for, and too small a margin of profit may well be a deterrent to good quality. It is conditions like this that must have inspired the doggerel:

The rain falls alike  
On the just and unjust fellow,  
But mostly on the just because  
The unjust has the just's umbrella.

These observations become clear in the hindsight of inspecting the record of what actually occurred. They might well have been clear beforehand, had one applied no more complicated statistics than that of the binomial distribution.

## TWO BASIC FAULTS IN THE SPECIFICATION

The analysis thus far, however, is wholly destructive. It neither points out the basic faults nor suggests their remedies. A little further use of engineering sense and statistical methods will accomplish these constructive ends.

In the critical examination of long records such as those encountered in large contracts one can make some important obser-

uations. One can clearly see, for example, that manufacturer A made a very good product, yet a moderate number of his good lots were rejected because of choice of acceptance criteria. On the other hand, the whole record shows that manufacturer B made a perfectly abominable product. He suffered a greater frequency of rejections than Mr. A, but he certainly managed to get many of his poor lots accepted before he went bankrupt or was barred from business, or perhaps he survived despite his lack of merit. All of A's lots should have been accepted: none of B's. One basic fault then lies in judging each lot on its own merits on a small sample basis which even in the case of homogeneous lots is obviously not an efficient basis for discrimination between good and bad lots.

If then, a large sample is economically prohibitive, one must turn to some other source of information. This information is, in general, abundantly available. It exists in the past performance of the manufacturer. It is a part and parcel of a very valuable asset which is generally called experience and expert engineering judgment. When data are costly and scant, no scrap of pertinent information should be allowed to escape; and statistics supplies a competent means (a) of determining the validity of the prior knowledge and (b) of mathematically combining the accumulated knowledge with that supplied by the sample itself.

The second basic fault is failure to make use of available knowledge other than that supplied by the sample, but the cogency of this criticism may not be apparent without a brief description of the mode by which the statistical method makes efficient use of this knowledge.

The technique of checking the validity of accumulated prior knowledge and of combining it with that supplied by the sample is a very common-sense rather than a technical procedure. It requires for its operation only three ideas: (1) the concept of statistical uniformity, (2) the operation of an elementary statistical technique, and (3) recognition of a simple empirical law.

Irrespective of homogeneity, one has almost certain (not just probable) knowledge of the product actually sampled. However, any inference drawn about the remainder of the lot inheres in some relationship, known or unknown, between sample and lot. Unless a product be homogeneous in the sense of being statistically uniform, inferences regarding the lot, based on the observation of samples taken from that lot, must be severely circumscribed,

since the reliability of sample quality as a witness of lot quality is unknown.

On the other hand if (a) statistical uniformity exists, and (b) one knows the general level of a quality characteristic, and (c) the measure of its variation, then statistical theory can predict the limits within which practically all samples of any size,  $n$ , should lie. One can obtain the knowledge of the general level of a quality characteristic and the measure of its variation from the accumulation of a large number of small samples. One can then predict the statistical limits of sampling fluctuation for practically all samples of size  $n$ . If a sample fall outside these limits, it appears that a state of statistical uniformity does not exist. If, however, no samples fall outside the limits, there is at least no reason for not believing the product to be statistically uniform. This inference is dependent upon an empirical law.

Extensive experience has shown that once a state of statistical uniformity is attained in a manufacturing process, the state of quality is inherent in the process itself, and cannot be changed without changing the process.

With these three ideas: viz., statistical uniformity, a statistical technique, and an empirical law, and with these three pieces of knowledge: viz., a state of statistical uniformity, the quality level, and the measure of sampling fluctuations, the small sample becomes a significant index of lot quality. The steps to significance are two. First, the state of statistical uniformity, level of quality, and measure of variation are inferred from the accumulation of samples. (This is the abundantly available information.) If statistical uniformity cannot be inferred from the data, then efficient discrimination between lots of good and bad quality by means of small samples is impossible. Second, if the quality be satisfactory, and if a subsequent sample fall within limits, there is no reason for believing that the satisfactory quality level has changed, but if it be not within limits, then the quality level appears to have changed from the known satisfactory level. (This inference comes from the information of the sample.)

These are essentially the principles used in Dr. Shewhart's Criterion I for quality control in manufacture. Concisely then, one may say, that whereas one cannot judge, except most approximately, the quality of the lot from the small sample, one can judge by statistical methods whether or not there is any reason for believing that the lot is of a different quality level from its predeces-



sors; and knowing the quality level of the predecessors from accumulated data, one can then infer the quality of the lot with considerable assurance. Hence, the second and corollary fault in the specification consists of not using hindsight as an intelligent guide to foresight.

## EVOLUTION OF A STANDARD OF QUALITY

In the light of these observations, let us consider the rôle of statistics in the development of a purchasing specification. One is in an ill position to write a specification unless one knows what one wants in the first place. Even if one knows what one wants, it does little good to specify it, unless there is at least some assurance that it can be met. Hence, that which is ideally desirable must be tempered by that which is practically obtainable. The only way to ascertain what is practically obtainable is through examination of records.<sup>1</sup>

From a long run of records one can readily determine the fraction defective, the average, the standard deviation, or other measures of the principal quality characteristics of the product which should be expected under good manufacturing practice. In my own case I generally consider every recorded test of the product under consideration and often the tests of allied products. Without statistical methods a fair consideration of such mass data would be almost impossible. Furthermore, the statistical method is essential as a means of inquiry into the homogeneity of the product, for without a state of statistical uniformity (as has just been pointed out), one is very limited in drawing conclusions regarding the lot based on the observation of samples.<sup>2</sup> If a homogeneous product of satisfactory quality level can be economically produced under good manufacturing practice, then one is in position to describe an economic standard of quality. Whereas the standard may or may not be the specification quality, it is nevertheless a desirable fundamental for writing the purchasing specification.

<sup>1</sup> When the product is new, resort must be made to the process described under "The Judgment of Statistical Control," Chap. I, pp. 43-46, & Chap. II, pp. 50-79, *Statistical Method*, W. A. Shewhart, Graduate School, Dept. of Agriculture, Washington, D. C., 1939.

<sup>2</sup> Chapter X, pp. 121-144, et. seq., *Economic Control of Quality of Manufactured Product*, W. A. Shewhart, D. Van Nostrand Co., New York, 1931.

## CONTRIBUTION OF STATISTICS TO THE DESIGN SPECIFICATION

It is impossible to specify completely a definite series of operations which will certainly detect that a very simple product is or is not of standard quality. Some of the quality characteristics are not capable of definite measurement, and after all there is a limit to detail. Hence, the specification of the intended quality must be limited to some chosen or principal quality characteristics of the product. This is termed the design specification, as it describes the design which it is intended that the specification shall obtain. Statistics makes an important contribution to the design specification as the intended or aimed-at values of the quality characteristics of the design, and their tolerance limits can be most concisely described by such statistical measures as the average and standard deviation. In surveillance work a quality description or design specification is prepared for each grade of ammunition of a given kind. Furthermore, it should be noted that a sound foundation for the statistics of the quality characteristics described in the design is provided by the investigation of the record which was made in the interest of the standard of quality.

## CONTRIBUTION OF STATISTICS TO THE ACCEPTANCE SPECIFICATION

The contribution of statistics to the acceptance specification is perhaps more marked. In the acceptance inspection, which is generally on a percentage basis, it is obviously impossible to determine with certainty that all of the product will meet the ideal of the design. Here again it is necessary to compromise with practicality and frankly admit that the logical mission of this specification is a mere statement of the quantity and kind of evidence which will be accepted as a satisfactory indication that the product will meet the standard of quality. It describes no indefinite ideals but confines itself to the delineation of a sequence of operations to be performed upon the product, the interpretation of the results of such operations, and a definition of the limits within which such interpreted results must lie in order for the product to be acceptable. In surveillance this procedure is, of course, the criterion by which lots are admitted to the respective grades described by the

design specifications for those grades. The objectivity of the statistical method makes it ideally adapted to the role of drastically limiting the need for interpretations of requirements. The necessity of interpretations and vague estimates usually opens the field for differences of opinion, whereas the statistical method swiftly produces the estimate of the measure under consideration in a manner superior to unaided judgment and without the bias always associated with personal consideration. This is a great step toward placing both purchaser and vendor on solid ground of mutual trust and understanding. Furthermore, in the light of the statistical technique the probability of accepting poorer than standard quality can be consciously minimized to any degree which appears to be economically desirable. One can do much to reduce the chance rejection of good lots and the chance acceptance of bad lots. Without an exhaustive investigation into the reasons for each step, which would be somewhat tedious, let us see one way (a number of others could be suggested) that a few modifications in principle can be made in the popular type of specification in order for it to have these manifold advantages.

### A STATISTICALLY SOUND ACCEPTANCE SPECIFICATION

*a.* From the first lot of a series  $k$  test samples of  $n$  shall be taken as nearly as practicable in order of manufacture, and the fraction defective,  $q$ , of each sample computed. The average of all  $q$ 's shall not exceed the specification value,  $q'$ , and no  $q$  shall depart from the average by more than three times the standard deviation of  $q$ .

*b.* From subsequent lots of a series, a test sample of  $n$  shall be taken at random and the fraction defective,  $q$ , computed. The observed  $q$  shall not depart from the average of all  $q$ 's by more than three times the standard deviation of  $q$  (based on all accumulated samples) nor shall it cause the average of all  $q$ 's to exceed the specification value,  $q'$ .

*c.* In event of a rejection, the next lot presented shall be considered a first lot of a new series.

Detailed steps in the specifications and provisions for retests have been omitted in the interest of focusing attention on statistical measures.

Part *a*, which involves a relatively large sample has two objectives. First, to subdivide the lot into sub-groups with respect to time in order that undue variation in the lot from part to part can

be detected. Second, to provide a large sample from which quality can be estimated with a high degree of precision, thereby providing effective discrimination between the manufacturer who has an established process in which the inherent quality level is good; i.e., of specification quality or better, and the manufacturer who is poor and who would ride as a parasite on the sampling fluctuations provided by a kindly chance. Part *a* makes relatively efficient discrimination on a small sample basis subsequently possible. It may be noted that the difference produced by the relatively large sample of  $kn$  is only one of degree, but it is nevertheless an important difference.

Part *b* is a step designed merely to detect a change in quality. It should be noted that it is necessary to take adverse action if there appears to be significant change in quality, even though the sample which is indicative of the change might be acceptable if the average quality level inherent in that manufacturer's process had happened to be lower but still within specifications. This situation must obtain since such an instance indicates either a change in average quality or a loss of statistical uniformity and hence results in an unknown situation, and one can no longer validly use the accumulated inspection data in predicting from sample to lot.

Part *c* is, of course, an avenue (though a somewhat painful one) for the manufacturer to reestablish himself.

These measures provide for an adjustment in acceptance criteria whereby the probability of rejection of a product of specification quality can be made very remote without prohibitive sampling. However, in attaining this end, the door has not been opened to the acceptance of poor quality, for: (a) the manufacturer has to qualify first on a large sample basis, and (b) if he should relax his quality after qualification, he might succeed in passing one or two poor lots, but the probability of several poor lots becomes increasingly remote and, once caught, he has to start all over with a new first lot of a series. The operation of these principles on accepted stocks not only shows their advantages in bold relief, but also shows the great gain that would be made if the influx of poor quality were checked at the time of its occurrence.

## SUMMARY

Statistical methods have been invaluable in evolving standards of quality for ammunition; for specifying grades based on these standards; and for writing the specifications for admission of lots

to these grades. In the field of procurement specifications their use has not been as extensive, because the need has not been felt so keenly. Nevertheless, some progress has been made. Statistical methods have proved to be a powerful tool in the critical examination of some ammunition specifications prior to final approval. Their use, either directly or indirectly, is almost essential in determining a reasonable and economic standard of quality through the method of comparing the quality desired with that which can be reasonably expected under good manufacturing practice. In like manner the statistical technique renders a valuable service in framing the acceptance specification. Through its use the quantity and kind of evidence which will be accepted as proof that the product will meet the standard of quality can be clearly expressed in a fair, unequivocal, and operationally verifiable way.

UNIVERSITY OF PENNSYLVANIA  
BICENTENNIAL CONFERENCE

The Relation of Statistical Quality  
Standards to Law and Legislation

*By*

ROSCOE POUND, PH.D., LL.M., LL.D., L.H.D., D.C.L., J.U.D.\*

LAW is a term used in three senses. The oldest meaning of the term among jurists refers to the body of authoritative precepts recognized or established as rules of conduct and of adjustment of relations among men, to be applied by the tribunals and administrative agencies of a politically organized society in determining controversies. Another meaning refers to what is perhaps better called the legal order—the regime of regulating conduct and adjusting relations by the orderly and systematic application of the force of a politically organized society. A third meaning refers to what is perhaps better called the judicial process, to which we must now add the administrative process—the process of determining the controversies involved in the regulation of conduct and adjustment of relations by applying to those controversies the authoritative precepts according to an authoritative technique. Two of those three meanings describe processes. But it was not till the present century that men thought of the legal order in terms of a process. Kant thought of it as a condition—a condition in which the free will of each was limited only by the like free will of all others according to a universal law. As to the third meaning, jurists of the last century simply ignored the processes by which the legal order is maintained, and concentrated on the precepts by applying which the condition (as they took the legal order to be) is realized. If, however, the nineteenth-century jurists overlooked the processes and saw only a condition brought about by precepts, the most insistent type of jurists of today ignore the precepts and apparatus and see only the processes, which, indeed, are two-thirds of the matter. For the legal order is a large-scale process, a specialized form of a still larger-scale process of social control, and the particular operations of the judicial process and

\* University Professor, formerly Dean of the Faculty of Law, Harvard University.

the administrative process are directed toward the ends of the large-scale processes and designed to achieve them.

Legislation is a process of framing the precepts by which the operations of the judicial and administrative processes are to be guided. Thus whether we are concerned with law or with law-making we are for the most part dealing with processes and operations, not, as men held in the last century, with conditions and rules. Here is the point of contact between the science of law, as a practical science, having to do with the attainment of ends by processes carried out with a minimum of friction and waste to plans which we call ideals, and the science of engineering, as another practical science having to do with attainment of ends by processes carried out with a minimum of friction and waste to plans and specifications usually in the form of blueprints. We must remember that "ideal" comes from a Greek word meaning picture, and is as much a picture as is the blueprint. Industrial engineering as a practical activity has a very significant analogy for the jurist. I have not hesitated to call the task of the jurist and lawmaker one of social engineering. Where the industrial engineer seeks to minimize, if he cannot eliminate, the element of chance and narrow the limits of expected deviation, so the juristic social engineer seeks to minimize, if he cannot eliminate, the personal element in the processes of the legal order and to keep expected deviation within closer bounds.

With reference to the judicial and administrative processes, we speak of this side of social engineering (for it is not the whole) as a quest of objectivity. This quest has been pursued by jurists since the Greek philosophers began to speculate about social control twenty-five centuries ago. To the Greeks, the phenomena of physical nature, the regularity and predictability of the return of the seasons, the phases of the moon, the succession of day and night, were helpful as suggesting an ideal of human conduct and exercise of the powers of social control equally regular and predictable. The phenomena of physical nature furnished a sort of blueprint. But they did not provide specifications for carrying out the plan. Indeed, for a long time jurists and practising lawyers did not get beyond the blueprints to the mode of controlling the operations so as to realize the blueprint plans as completely as possible.

As Dr. Shewhart puts it, it is not merely how we aim but how we hit that counts. General Schofield tells a story of a volunteer

and a regular soldier who had fought side by side at the battle of Wilson's Creek. The volunteer's cartridge box was empty, and he asked the regular how many rounds he had fired. The regular looked at his cartridge box and answered "nineteen." The volunteer explained proudly that he had fired forty. "Yes," said the regular, "how many rebels did you hit?" The answer came, "I don't know; how many did *you* hit?" Said the regular, "I reckon just about nineteen." Control of operations so as to make results conform as closely as possible to the ends sought is much the same problem in a factory, in an army, or in a legal order.

For a time, objectivity was sought by attributing divine authority to the body of legal precepts, assuming that the authority behind them would sufficiently insure that they would be applied and their inherent objectivity would take care of the process of application. Later, a basis of authority was found in custom on the theory originally, I suppose, that long-continued usage showed that the gods approved of the customary precepts, and so it was safe to apply them and unsafe to risk others which the gods had not approved. Still later, the authority of custom was referred to the superior wisdom of an ideal past, and then authority for a time was held self-sufficient or was bolstered up by an authoritative philosophy. Next, for a time, reason was put as the basis of the precepts and of their application, and the analogy of geometry was relied on to show how applications could be deduced infallibly from authoritatively prescribed or discovered premises. In the last century, metaphysics and a metaphysical conception of history as the unfolding of an idea, and so of legal history as the unfolding of an idea of right or idea of freedom, were taken to show the path to objectivity in the legal order; and in the second half of the century such ideas began to give way to a positivist conception of laws of social development analogous to the laws of physical nature as they were then understood. In all these successive theories of law, objectivity of application is assumed to be involved in the nature of the legal precept, although the classical natural law, taken to rest upon reason, did think of application guided by reason as well as precepts expressing reason.

In the last quarter of the nineteenth century, Jhering with his theory of end or purpose in law and social-utilitarian philosophy of law, not as something self-developing but as designedly framed and applied for social ends, started a new movement in jurisprudence which is still felt, especially in America. In this theory of



law the process, and direction of the process consciously toward conscious purposes, was stressed, and legislation and administration, which the historical jurists of the last century considered futile or negligible, found important places in the science of law. It is true this turning of attention to the processes, without more theory of guidance of the processes or measuring the conformity of their results to the assumed ends than acceptance, express or tacit, of the older idea that the precept somehow involved in itself the guarantee of objective application, led presently to a reaction from the whole idea of objectivity and assumption that in the nature of things it was not attainable. Different varieties of self-styled realism—for their use of the name “realism” is often more a boast than a description—and phenomenalism have grown up which urge on one ground or another—economic determinism or Freudian psychology, or Neo-Kantian relativism or all embracing skepticism—that all we have been trying to do and have believed we were doing in law since the classical Roman jurists, whose books have been quarries ever since, is mere illusion or self-deception or superstition or pious fraud. A recent reaction from the extreme views of realists and phenomenologists has led to a revival of scholasticism, of the official philosophy of the Catholic church, the philosophy of St. Thomas Aquinas, and neo-scholastics in many lands are turning back to a basis of philosophical-theological authority.

Lawmaking and application of law have been guided by received ideals which are as much a part of the law as the precepts developed, interpreted and applied to their pattern. Indeed, objectivity in the phenomena of the legal order is itself an ideal. It is always hard to attain ideals, and one may concede that no more than an approximation sufficiently close for practical purposes may be expected. Moreover, twentieth-century psychology and economics make the ideals of social control and so of the legal order appear much more difficult of attainment than we took them to be. Hence it is now the fashion to give up the ideals, to give up the age-long quest of an objective process, and yield to or even get behind the revival of political absolutism. But to give up in this way is to give up half of civilization.

It will have been seen that at the moment the science of law is very much at large. There is no agreement about fundamentals or even whether there are fundamentals. But a type of realist has been insisting that somehow we may build better on a basis of

statistics. Little has come of this, however, and it will be instructive to see why, before I turn to the light thrown upon the possibility of control of operations of lawmaking, and application of law through the judicial process, by Dr. Shewhart's discussion of quality control.

For one thing, the realists who have been urging that in some way or other properly kept and properly organized statistics of the operations of the judicial process would go far to help toward solution of problems of jurisprudence have had a faith, curious in those who are in general so skeptical of everything but their own skepticism, that a mass of figures have an intrinsic objective validity, and that a bit of juristic writing is scientific if only it has a sufficient number of tables of figures. Matters which have been well known from long and general experience are regarded as requiring to be proved by such tables although the tables may very likely be constructed on the assumption of what they are called upon to prove.

In the next place, the unreliability of our official statistics, national and state, so far at least as they have to do with the operations of the legal order, has stood in the way of any effective use of them. Indeed, it requires very little experience of trying to use them to convince one of the truth of Bill Nye's saying that figures won't lie but liars will figure.

Bureaus seeking to make a showing in order to obtain increased appropriations, clerks set to the task of preparing reports and proceeding to pad them with statistics for statistics' sake, and well-meaning blunderers preparing tables with no clear idea of what is or may be significant in the materials before them, have contributed to discourage the conscientious judge or advocate or teacher who would like to know the facts underlying current assumptions as to American justice. There have come to be abundant official tables and masses of figures. But they are seldom useful except for the purpose for which they were prepared, namely, to make a showing. On inquiry they are likely to prove inaccurate, uncritical, contradictory, and not comparable with each other. For example, some statistics as to the prevalence of crime in different parts of the country published a few years ago showed a much greater amount of larceny in a rural agricultural community than in one of the old settled states, chiefly urban and industrial, and with a shifting population. The compiler had not noticed that in the former state by statute larceny had been ex-

tended to embrace embezzlement, obtaining money or property by false pretenses, and receiving stolen goods, while in the latter the narrow definition of the English common law still obtained and only a taking from the possession of another without his consent and with intent to convert to one's own use was called larceny. Again, not many years ago different federal bureaus dealing with different aspects of federal enforcement of a certain statute published tables in which the same subject was treated on different bases so that the figures could not be made to agree. In some jurisdictions, tables have been drawn up so carelessly or so perfunctorily that the numbers reported by one as sent to penal institutions and the other reporting the number received by the same institutions did not agree. Certainly it was not safe to infer from the serious discrepancy that a large number of those sent to prisons had escaped in transit. In one case within my personal knowledge where tables published by different bureaus in the same jurisdiction, which should have agreed in the totals, were quite out of accord, when attention was called to the matter and revision was promised and supposed to have been made, it turned out that the revision had brought the totals into accord by a simple process of wrong addition. A mean was struck between the discrepant totals and the arbitrary result was put at the foot of the respective columns. Nothing could be more discouraging than experience with statistics published by state bureaus and departments in this country.

Of late, judicial councils in a number of states have been putting out some reliable statistics of the work of the courts, carefully compiled to an intelligent plan and directed to certain ends and so useful for purposes clearly manifest. But even when statistics of that sort become general, the difficulty will remain that the realists who have been urging a science of law based on statistics have expected too much of them. We can't expect to use them to solve the fundamental problems of jurisprudence. They can't give us a measure of values of competing claims or a criterion of justice or a theory of what we are seeking to bring about by means of the legal order. But it does not follow that we have no use for statistics. On the contrary, the stress which we now put upon the legal order as a process and upon the judicial process as a significant meaning of the term law, and along with the administrative process as entitled to a place in the front rank in the science of law, things which the realists have been prominent in urging, in-

dicates where statistical method is to be made use of. We must learn how to use statistics to control the quality of the output of the operations by which the legal order is maintained and carried on.

As long ago as 1823, Jeremy Bentham, the founder of the science of legislation in the English-speaking world, urged a ministry of justice as the means of improving legislative lawmaking. Since Bentham it has been urged by many of the leaders in movements for the improvement of the administration of justice, notably in recent years in the United States by the late Mr. Justice Cardozo. It has been urged also that such an institution could do much to improve the quality of the output of the judicial process. But those who have urged this necessary agency of improvement have always been asked how we may be sure that a ministry of justice will be able to bring about the quality of lawmaking and of application of law which we desire. Answer has not been easy. It is now commonly asserted that judgments of quality can never be more than subjective opinions of no scientific value. When it is urged that a ministry of justice will give us properly kept and intelligently organized statistics of the operations of the legal order, answer is made that we have still to develop adequate technique and well worked out principles before we may be assured that our ministry of justice, if such things can be measured or controlled at all, will be able to pass objectively upon the quality of the legislative and judicial product and enable us continuously to improve it.

When Dr. Shewhart was good enough to send me his illuminating paper "Some Aspects of Quality Control," I saw at once its bearing on this question of a ministry of justice. Quality control in the legislative and judicial and administrative processes was what jurists had been looking for. "Verifiable operational quality" and "objective judgment of quality" were phrases of familiar sound putting from the standpoint of the industrial engineer problems which had long troubled the social engineer in the fields of law and legislation. Very likely my enthusiastic response to Dr. Shewhart was the *causa causans* of my appearance here in the capacity of an engineer.

Let me endeavor to point out some of the possibilities of an engineering theory of quality control to the work of a ministry of justice.

It is significant at the outset that quality is not defined as something absolute but as related to a process—a succession of perceivable qualia, i.e., repetitions of attributes, associated with a previ-

ously conceived or specified set of operations. In the science of law we have to think of quality not as absolutely determined by abstract reasoning or by logical deduction from some unchallengeable metaphysically given datum, as in the legal philosophy of the past, but as relative to processes of making or finding precepts or of determination of controversies or to continuous demands which are analogous to processes. As to the lawmaking and determining processes, the succession of qualia we seek are in the same way associated with a previously conceived and specified set of operations. Even in the case of demands or claims, which make one of the most difficult problems of our science, there is a suggestive analogy in the types of quality laid down for the industrial engineer. If we are considering the quality of a demand or claim, type one has to do with its character independent of other demands or claims, e.g., the intensity of the demand. Type two has to do with its character as one of a group or body of claims or demands, e.g., individual, public, or social. Type three has to do with its character of being asserted generally, widely, by many persons or otherwise. Can these things be determined objectively? One should ask this in comparison with the practical method of the past, namely, determination by experience of what will satisfy the most of the whole scheme of demands or claims with the least friction and waste. But more can be made of the method of quality control with reference to the processes of the legal order.

What can we derive from Dr. Shewhart's theory and technique of quality control? In the judgment of quality, we are told, the first step is legislative in character; specification, to be compared with legislative or traditional fixing of the ends to be attained by a precept. The next step is one of judging whether a determination when made by applying the precept is likely to meet the specifications and whether there is any evidence that the specifications should be changed so as better to adapt them to realizing the ends prescribed. For the judgment of quality we need not only specifications of form but also some definite indication of the degree of assurance that we are to have, in rendering the judgment, that the quality of a particular application of a precept or judicial or administrative determination will meet its associated specification. There must be an element of uncertainty in fixing the specifications. The function of fixing them is legislative and is tempered in practice by a judgment based on experience as to the attainableness of any specified goal. It will be seen here that I have put one

of the chief tasks of a ministry of justice in almost the very language of engineering quality control. Again, the task of judging of quality involves seeing to it that any failures to hit the mark prescribed are analyzed in such a way as to indicate whether or not it is desirable to modify the specifications. This is as much the job of a ministry of justice as it is of an industrial engineer. In the one case, too, as in the other it is important to determine how much information is to be collected and how it is to be collected. Incomplete specifications have to be interpreted and amended so as to make them more complete; the evidence as to defects of specification has to be interpreted; allowance has to be made for counterbalancing effects of other specifications and a running record kept of the evidence. Thus evidence may be found of assignable causes of deviation and the deviations may be traced back to lack of control of specified qualities or to the insufficiency of specified characteristics. This could have been written of a ministry of justice or judicial council as well as of an industrial engineer.

So, too, of the method of making a quality specification objectively verifiable. First, the qualia to be produced are to be specified (and herein lies much of the quarrel between lawyers and the ambitious young advocates of administrative absolutism), and, second, the detailed technique of verification is to be prescribed. There is always an uncertainty that any set of operationally verifiable specifications constitutes the ideal set, and so the jurist has no occasion to feel that he cannot achieve a reasonable practical approximation any more than the engineer.

Preparation for legislation under the influence of natural-law ideas consisted in putting different proposed precepts and those already promulgated in other jurisdictions side by side and choosing the one that, subjected to the scrutiny of reason, seemed most abstractly just. The program of sociological jurists calls for study of the actual social effects of legal institutions, legal precepts, and legal doctrines. Rationalist natural law insisted exclusively upon reason. Positivism insisted exclusively upon experience. What we seek is experience developed by reason and reason tempered by experience. But experience is not something we can grasp and organize by offhand individual recollection and observation. The merit of the theory of quality-control for the jurist's purposes is that it directs statistics toward definite preconceived ends instead of expecting statistics gathered to no clear purpose to reveal the something for which they are to be used.

It is not that a ministry of justice or a judicial council can take Dr. Shewhart's pamphlets and use them as manuals. But if we cannot use his precise methods for our purposes, for which they were not designed, we can use them to indicate to us how to go about setting up an objective method of our own. We can perceive what to require of a method and what we may expect it to do. We can be helped to do intelligently what we have been doing by a halting and sometimes blind empiricism.









W  
3073